TRACKING AND VERIFICATION OF TROPICAL CYCLONE DEVELOPMENT IN GLOBAL ENSEMBLE PREDICTION SYSTEMS: EVALUATIONS DURING RECENT FIELD PROGRAMS

by

Andrew David Snyder

A thesis submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Master of Science

Department of Atmospheric Sciences

The University of Utah

December 2009

THE UNIVERSITY OF UTAH GRADUATE SCHOOL

SUPERVISORY COMMITTEE APPROVAL

of a thesis submitted by

Andrew David Snyder

This thesis has been read by each member of the following supervisory committee and by majority vote has been found to be satisfactory.

Pu

S eenburgh

THE UNIVERSITY OF UTAH GRADUATE SCHOOL

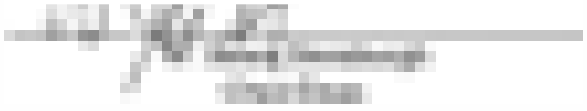
FINAL READING APPROVAL

To the Graduate Council of the University of Utah:

I have read the thesis of Andrew David Snyder in its final form and have found that (1) its format, citations, and bibliographic style are consistent and acceptable; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the supervisory committee and is ready for submission to The Graduate School.

Chair: Supervisory Committee

Approved for the Major Department



Approved for the Graduate Council

ABSTRACT

This study evaluated two global ensemble forecast systems from the National Centers for Environmental Prediction (NCEP) and the Naval Research Laboratory (NRL) in their performance of predicting the genesis and evolution of tropical cyclones. The skill of the ensembles was assessed using cases studies from recent field programs. For each model, the case studies consisted of five named tropical cyclones and two unnamed nondeveloping tropical systems. The overall skill of ensemble forecasts, ensemble means and spreads, and probabilities were verified.

Various tracking methods with different variables and vertical levels were compared for effectiveness. A manual tracking method with three variables (vorticity maxima, geopotential height and circulation) at the 850 hPa level was adopted to track tropical cyclones in ensemble forecasts for both their pregenesis and postgenesis phases.

In the NCEP ensemble forecasts, the overall forecast skill of ensemble tracks and intensity in the storm's pregenesis phase was relatively low. The forecast uncertainty generally decreased with the reduction of forecast lead time, while the short-range forecasts tended to be more accurate. In contrast, the skill of ensemble track forecasts significantly improved for matured storms, possibly because of the implementation of the NCEP storm relocation scheme. Although the high resolution deterministic GFS forecasts were equal or better than the ensemble forecasts in almost half of the cases, the ensemble

forecasts added value in predicting the probability of the tropical cyclone genesis. For example, the ensemble spread could be a good indication of the forecast track errors.

Two versions of the NOGAPS ensemble are compared: with and without the addition of stochastic convection. The forecast skill of tropical cyclone tracks and development from the NOGAPS ensemble without stochastic convection was relatively low, although it provided additional information over the deterministic control. Ensemble track spreads were small and underdispersive in the control ensemble. Including stochastic convection enlarged the ensemble track spread and enhanced genesis predictability. The ensemble spread also showed a better correlation to the forecast track errors, especially in the postgenesis phase. However, further evaluation showed that weaker tropical systems could be over-intensified in the ensemble with stochastic convection.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

I would first like to thank the members of my committee for their advice, hard work, and encouragement to make this thesis possible. My adviser, Dr. Zhaoxia Pu, has been immensely helpful over the past two years. She facilitated this research through advice, direction, help with problems and coding, and always making sure I was on the right path. Dr. Jim Steenburgh has shared many helpful thoughts and suggestions on both this research and other graduate school issues. Dr. Carolyn Reynolds has been an excellent collaborator, especially with the NOGAPS research, by offering support and new ideas. She also generously made the NOGAPS ensemble data available to me. Thanks to the faculty of the University of Utah Atmospheric Sciences Department, who provided me with a solid graduate school education and challenged my thinking. I would also like to thank all of my friends, both near and far, who have supported me and provided excellent stress relief. Most importantly, my amazing family has been by my side to support and push me to excellence.

A special thanks to J. H. Zhao, a visiting scholar from China hosted by Dr. Zhaoxia Pu, for initiating part of tracking programs used in this study. Thanks also to Yuejian Zhu from NCEP/EMC for his help and comments on the NCEP global ensemble products and to Dr. Richard Wobus for his help in transferring the NCEP ensemble data to the University of Utah. This study was supported by the Office of Naval Research through the TCS-08 program.

CHAPTER 1

INTRODUCTION

Numerical weather prediction (NWP) has continuously improved since widespread use began half a century ago. However, forecast errors exist due to uncertainties in model initial conditions and imperfect physical parameterization schemes. Specifically, forecasts of the genesis and evolution of tropical cyclones remain a great challenge for NWP due to a lack of in-situ observations over vast ocean areas. Advances in computer science and computer power have, however, made it possible to use ensemble forecasts to account for uncertainty in model initial conditions. Since the early 1990s, ensemble forecasting has become operational in many major NWP centers around the world (e.g., Toth and Kalnay 1997; Buizza et al. 2005; Wei et al. 2008; Reynolds et al. 2008).

During hurricane season, tropical cyclone forecasting is a high priority in many of these operational centers. Owing to the serious economic and social impact tropical cyclones cause, it is important to predict their genesis and evolution with enough lead time and accuracy. In the last two decades, along with advancements in numerical modeling and data assimilation, track forecasts for mature tropical storms have improved significantly. However, forecasting tropical cyclone genesis and intensity changes remain challenging problems (Rogers et al. 2006). Aberson (2001) found that ensemble techniques were inadequate at estimating the forecast reliability. He suggested that these

shortcomings could be rectified only when models were improved or perturbation techniques became more sophisticated. Rogers et al. (2006) concluded that intensity forecasts for tropical cyclones have been slow to improve due to lack of data, modeling limitations, and gaps in understanding of physical processes. More specifically, operational models have historically struggled with the prediction of genesis and often produce spurious vortices (Beven 1999). Several studies have, however, demonstrated that as operational models become more complex, they do have predictive skill in the tropics. For example, Rennick (1999) compared the skill of the Geophysical Fluid Dynamics Laboratory's (GFDL) hurricane prediction system with the Navy's Operational Global Atmospheric Prediction system (NOGAPS). Errors were smaller for stronger systems, while track predictions were more reliable than intensity forecasts. In addition, Pasch et al. (2002) found that improvements to the NCEP Global Forecast System (GFS) have enabled greater skill in the prediction of tropical cyclone formation. However, genesis prediction and differentiation between developing and nondeveloping systems continues to be a challenging problem, and thus remains a large area of research (e.g., McBride and Zehr 1981; Perrone and Lowe 1986; Hennon and Hobgood 2003; Kerns et al. 2008).

The advent of ensemble forecasting has added a new dimension to tropical cyclone prediction. Instead of a single deterministic forecast, a suite of forecasts adds a probabilistic component to the forecast, thus helping estimate forecast uncertainty. Model spread for cyclone tracks as well as atmospheric conditions has led to increased forecast confidence and skill for genesis, cyclone tracks and landfall effects. So far, tropical forecasters have largely taken a consensus (multimodel) approach when utilizing

ensemble forecasting. This approach is primarily preferred because studies have shown that forecast errors for individual members of single model ensembles can be large. The ensemble mean errors in single model ensembles can also be much larger than those of multimodel ensembles (Goerss and Reynolds 2008). In addition, single model ensembles usually have less error independence amongst the members. Commonly, all (or some) models available to forecasters (e.g., these from different operational centers) are utilized. In this manner, the multimodel approach is advantageous since nonnumerical and statistical models are included in the ensemble. Another approach statistically combines different models that were either designed for or have a history of performing well in the tropics, such as those developed by Goerss (2000), Krishnamurti et al. (2000), and Weber (2003).

In many operational centers, such as the U. S. National Centers for Environmental Prediction (NCEP) and the European Centre for Medium-range Weather Forecasting (ECMWF), ensemble forecasts are produced from a single operational model, usually at a coarser resolution in consideration of computational expense. With a single model, the ensemble is produced by using perturbed initial conditions (although the ECMWF also perturbs model integration). However, almost all perturbation methods used for current operational models (e.g., breeding vectors, singular vectors) are based on mid-latitude variability and may not be very suitable for tropical ensemble forecasting. However, the ECMWF currently uses tropical singular vectors. In response to the lack of tropics-specific ensembles, Zhang and Krishnamurti (1997) developed an empirical orthogonal function based perturbation method for the tropics. It proved to be skillful in a reforecast of Typhoon Winnie (Mackey and Krishnamurti 2001). Cheung and Chan (2001) also

experimented with tropical ensemble forecasting using a barotropic model, with some success. Otherwise, there has been little work done to evaluate the skill of operational ensembles in the tropics. Recent studies from Aberson et al. (1998), Aberson (1999), and Marchok (2002) showed that the NCEP's global forecasting model and ensemble forecasting system were skillful for tropical cyclone track prediction, but little attention was focused on the skill of the ensemble forecasts for tropic cyclone genesis. Goerss and Reynolds (2008; discussed more in Chapter 3) have performed a similar evaluation of track forecasts in the NOGAPS ensemble system, but their study focused solely on the postgenesis phase of cyclones.

Theoretically, since ensemble forecasts provide probabilistic information on both vortex spin-up and ambient environmental conditions, they could prove to be useful in predicting tropical cyclone genesis. Therefore, the purpose of this study is to evaluate the NCEP and NRL global ensemble forecast systems for their ability to predict the genesis of tropical cyclones. The primary focus will be tracking the development of tropical storms and then evaluating how well the ensemble systems predict tropical cyclones from their early pregenesis phase to a developed cyclone. Predictability will be assessed in both a pregenesis and postgenesis environment. Track errors, mean, and spread will also be evaluated. For the NRL NOGAPS global ensemble, two different ensemble methods, with and without stochastic physics, will be compared for their ability to predict tropical cyclone genesis and evolution.

Detailed evaluations of NCEP global ensemble, as well as the effectiveness of tropical cyclone tracking methods, are described in Chapter 2, focusing on cases that occurred during the NASA African Monsoon Multidisciplinary Analyses (NAMMA)

between Aug and Sep 2006. Chapter 3 addresses the comparison of the two different ensemble methods within the NOGAPS global ensemble, using selected cyclones from the THORPEX Pacific Asian Regional Campaign (T-PARC) during Aug and Sep 2008 is addressed in Chapter 3. Concluding remarks are made in Chapter 4.

CHAPTER 2

NCEP ENSEMBLE PREDICTION OF TROPICAL CYCLONE DEVELOPMENT: EVALUATION DURING NAMMA

2.1 Introduction

In this chapter, an evaluation of five tropical cyclones and two nondeveloped tropical waves forecast by the NCEP Global Ensemble Forecast System will be discussed. All seven cases occurred concurrently with the NAMMA field experiment between Aug and Sep 2006 (Zipser et al. 2009). The mission of this field campaign was to characterize the structure and evolution of African easterly waves and mesoscale convective systems and their impacts on regional water and energy budgets. The development and evolution of tropical cyclones were studied, as well. In addition, impacts of the Saharan Air Layer and other aerosols were investigated.

The five named cyclones include Tropical Storm Debby, and Hurricanes Ernesto, Florence, Gordon, and Helene. Debby, Florence, and Helene were studied and sampled during NAMMA. The two nondeveloping waves (unnamed and numbered as Waves 3 and 6 in sequential order, following Zawislak and Zipser 2009, in press) also traversed the eastern Atlantic during this time period, and were briefly investigated for comparative purposes. Fig. 1 shows the tracks of the five named tropical cyclones studied in this thesis.

The chapter is organized as follows. Section 2.2 briefly describes the NCEP ensemble system and the data used in this study. Tracking methods and criteria are

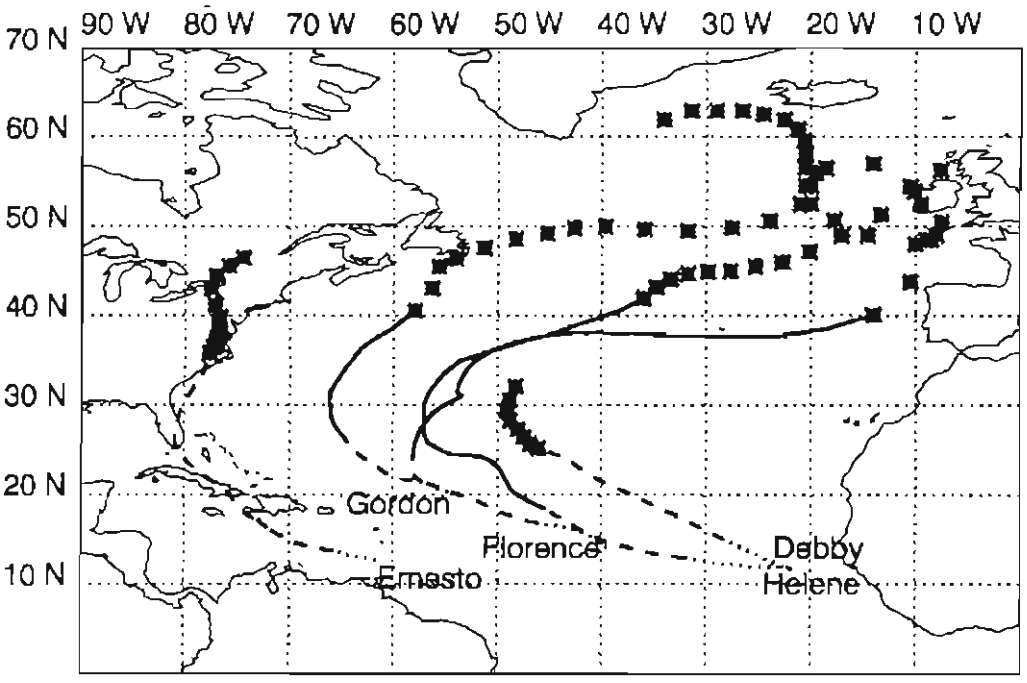


Figure 1: National Hurricane Center (NHC) best track for the five named cyclones investigated. Dotted line represents a tropical depression, dashed line represents a tropical storm, solid line represents a hurricane, and cross represents a remnant low.

evaluated and defined in Section 2.3. In Section 2.4, the results are summarized from each case study, and Section 2.5 contains an overall evaluation. Section 2.6 presents concluding remarks.

2.2 Brief description of the NCEP ensemble system and the data used in this study

The ensemble forecasts used in this study were obtained from NCEP. There are 14 ensemble members at T126L28 resolution available for each set of ensemble forecasts. Initial perturbations are generated using the ensemble transform breeding method as described by Wei et al. (2008). A hurricane vortex relocation method is implemented (Liu et al., 2006) to locate the cyclone in the correct position in each of the members. This procedure can be summarized in the following steps. A spectral filter is used to separate the total wind field into "basic" and "disturbance" fields. Then, the hurricane vortex is located in the disturbance field. The model's vortex is then separated from the nonhurricane component in the disturbance field. Next, the basic wind field and nonhurricane component are combined into the environment wind field. The extracted vortex is relocated to the NHC's official position. If the vortex is too weak, bogus observations are added in the model analysis. Finally, the data assimilation scheme uses the revised guess field plus available observations to produce a final analysis. In addition, the cyclone intensity is also perturbed with 5% of the magnitude of the wind speed. For brevity, the evaluation presented here concentrates on the forecasts starting at 0000 UTC each day and proceeds to 120 h.

The operational Global Forecast System (GFS) forecasts, produced at T382L64 resolution, were also utilized for comparison. This baroclinic-dynamical model was not

developed specifically to track and predict tropical cyclones, which is compensated by utilization of vortex relocation. The addition of momentum mixing into the cumulus parameterization scheme reduced tropical storm false alarms (Pasch et al. 2002). While forecasts are available in 3-hour time steps, only the 6-hourly forecasts are used, and likewise until the 120 h forecast.

2.3 Evaluation of the tracking methods

The primary focus of this study is to track the tropical systems in both the pregenesis and postgenesis phases. The *genesis time* in this study is defined as the time when the system was designated by the National Hurricane Center (NHC) as a tropical depression in the best track data. This time was chosen as a reference point because it serves as a clear dividing line between an unorganized and organized tropical low pressure system. In addition, each tropical depression took a different amount of time to reach its tropical storm status, which would make direct comparison difficult if the time when a system reaches the tropical storm intensity was defined to be the genesis time. To adequately assess pregenesis and postgenesis forecasts, cyclones are tracked from approximately three days before they were named to 2 days after they have been designated a tropical depression. All track comparisons to the "actual cyclone" are made with the NHC's official best track data (retrieved from http://www.nhc.noaa.gov/2006atlan.shtml).

Tracking the genesis of tropic cyclones often proved to be challenging since the center of the system was not always obvious, especially in the pregenesis phase. Therefore, an effective tracking method has been developed to accurately represent the track of the cyclone.

2.3.1 Review of previous tracking methods

Three types of tracking methods for tropical disturbances can be found in the literature: manual, automated, and statistical. Manual tracking involves human analysis of data utilizing a number of different variables. Several studies have tracked cloud clusters using satellite data (e.g., McBride and Zehr 1981; Perrone and Lowe 1986; Hennon and Hobgood 2003). One advantage of this method is the frequency of the data; however, cloud clusters do not always correspond to the center of the tropical disturbance. Another popular manual tracking method employs a large-scale analysis or reanalysis and uses the low level vorticity and the meridional wind component (to identify passage of the wave trough). Hovemoeller diagrams have also been used to aid in tracking (e.g., Fink et al. 2004; Chen 2006; Kerns et al. 2008). Reed et al. (1988) found that tracking waves consistently was difficult, so they preferred to use vorticity. However, they also noted that weak systems could have multiple vorticity centers which made the wave track ambiguous. Kerns et al. (2008) reiterated these difficulties along with complications from spurious small-scale vorticity "bullseyes" and vorticity maxima that formed and dissipated within the same wave. To counteract these problems, they instituted a Lanczos bandpass time filter and Gaussian spatial smoothing so that only consistently strong vorticity maxima appeared in the analysis.

Thorncroft and Hodges (2001) developed an automated tracking system that used a cost function to match features between time steps in the analysis. They noted that multiple vorticity centers and wind shifts from squall lines caused difficulties with their algorithm. Operational centers have also instituted automatic cyclone track algorithms for named cyclones, such as used by NCEP (Marchok 2002). Several studies (e.g., Burpee

1972; Albignat and Reed 1980; Pytharoulis and Thorncroft 1999) have utilized statistical approaches including power spectra, composite charts, and the kinetic energy of horizontal winds. However, such analyses suffer from excessive smoothing of the fine-scale structure of the tropical disturbances. Both automated and statistical methods also can be limited by the criteria and thresholds used.

As discussed above, most of the previous tracking methodologies have been developed for tropical cyclone analysis and not model forecasts. Compounding the problem, many existing cyclone tracking methods are designed only for the postgenesis phase. Considering the limitation from both automated and statistical methods, all cases for this study are manually tracked on a plan-view map of model output. By performing manual tracking, attention could be focused on the details of each tropical system, which can be important in the pregenesis stage. Meanwhile, since there is not a standard variable (or parameter) or height level that has been identified to be most effective tracking tropical cyclone development, an evaluation has been completed to determine a reasonable way to track tropical systems.

2.3.2 Tracking criteria

Tropical waves, depressions and cyclones are often characterized by geopotential height minima, vorticity maxima, and cyclonic circulations. How well these characteristics are captured by numerical models strongly depends on the model resolution. In order to make a consistent definition of tropical cyclone genesis in the ensemble resolution, a statistical analysis was conducted using the NCEP GFS final analysis (FNL) at the same grid spacing. By analyzing the magnitude and structure of the height, vorticity and circulation fields in the GFS analysis at the time closest to the actual

genesis of the five cases being studied, it is apparent that most cyclones were represented by multiple closed height contours (at a 10 m interval) within 5 degrees of the low pressure center, cyclonic circulation in the wind field, and a relative vorticity maxima greater than 7.5×10^{-5} s^{-1}.

Therefore, each ensemble forecast initialized in the pregenesis phase was given one of four designations. When all three of the above criteria were met in a forecast, it was noted that the forecast predicted the cyclone's "genesis." If one or two of those conditions were depicted in the ensemble forecasts, then the forecasted system was labeled as "vortex-like." "Dissipation" was defined to mean that the parameters for genesis existed but did not persist for longer than 48 h. If none of the criteria were met at any time in the forecast period, the forecast was designated as "nondevelopment." These criteria are mostly consistent with those used by Cheung and Elsberry (2002), who tracked tropical cyclone formation over the Western North Pacific with the Navy Operational Global Atmospheric Prediction System (NOGAPS) model.

2.3.3 Tracking variables

In order to evaluate the different parameters for their effectiveness in tracking the cyclone development and genesis, geopotential height, vorticity, and wind vectors were tested based on their importance in depicting storm structure. Thermal variables were not considered because independent investigation revealed that the warm core structure was not always evident at this grid spacing. Each case was tracked independently with each variable, and the tracks were then compared with the available NHC best track data.

In a majority of the cases, there was very little difference in the tracks using the three different variables. Fig. 2 shows the overall averaged track errors with different

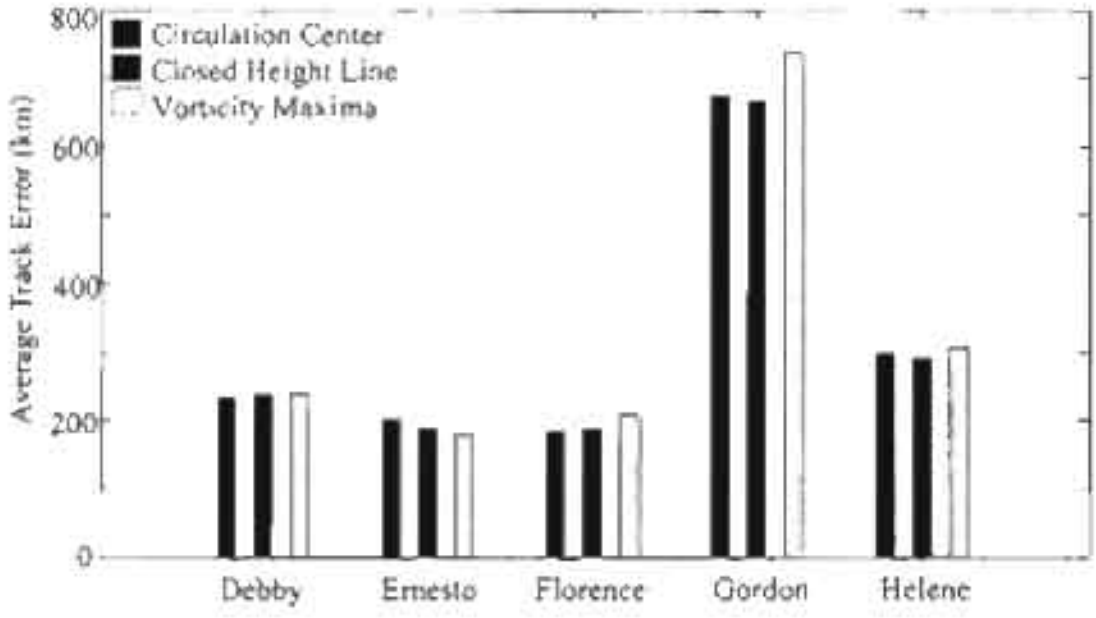


Figure 2: Distance error (in km) for each tracking variable, averaged over all five named cases for both pre and post genesis phases for 1-5 day ensemble forecasts over all members from all different lead times (1-3 days before genesis and 0-2 days after genesis).

variables over all cases for both pregenesis and postgenesis phases for the one- to five-day ensemble forecasts of all members. Overall, the averaged track errors for each case were very similar regardless of which variable is used for the tracking. However, the most drastic differences occurred for very weak systems (e.g., in the wave phase) or for complex systems (e.g., the first stages of Florence, discussed in Section 2.4). Fig. 3 shows a typical example with the tracks of Florence using different variables. Tracking started at 0000 UTC 1 Sep 2006, 66 h before the system became a named tropical depression. It is apparent that the discrepancies among the tracks using different variables were large when the system was in its weak phase. However, after the system was forecast to organize, all variables agreed in cyclone track accuracy.

In addition, the relative importance of each variable was case dependent. Fig. 4 shows plots of vorticity, wind, and geopotential height from actual tracking sequences. These selections were examples of cases which demonstrate the need for multiple tracking variables. For instance, in the pregenesis phase (Fig. 4a), the center of the vorticity maximum was generally most useful since the system was often broad and erratic. After genesis (Fig. 4b), the center of the closed height contours best represented the cyclone. In some cases, the system had both a broad area of vorticity and closed height contours (Fig. 4c), in which case wind vectors helped to identify the center of the system. However, the wind field was generally the least useful due to the relatively coarse resolution of the model. Additionally, weak and deteriorating waves often had ambiguous centers (Fig. 4d). In these cases, the manual tracking method based on the judgment of all three variables was advantageous in picking the center of the feature consistently.

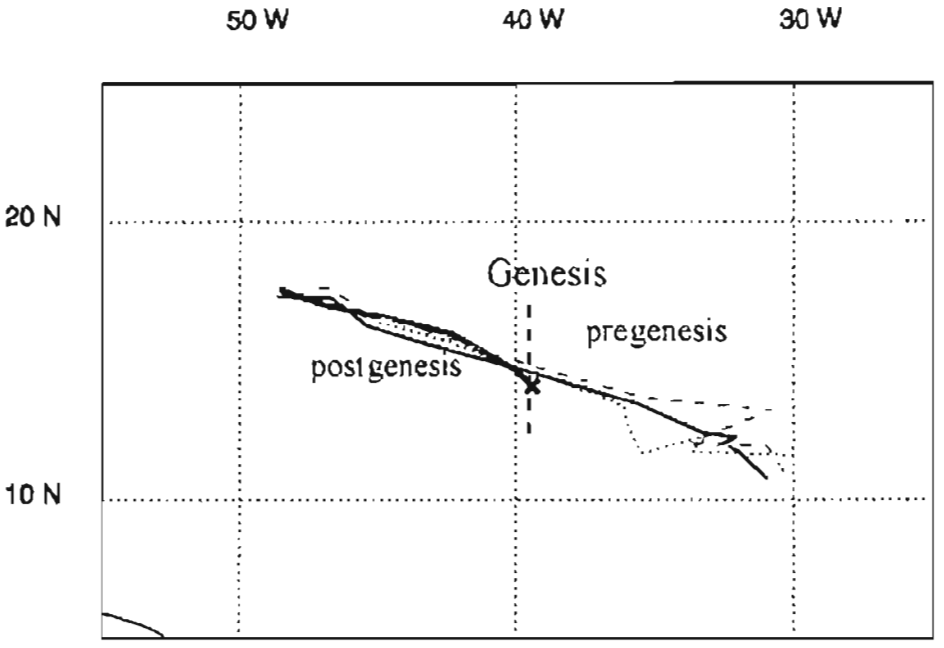


Figure 3: An example of the 120-h ensemble mean track differences among three variables used, from 66 h before Florence was designated a tropical depression (0000 UTC 01 Sep 2006). The solid line denotes the track following the center of the circulation, the dashed line presents the track following the vorticity maxima, and the dotted line links the centers of closed height lines.

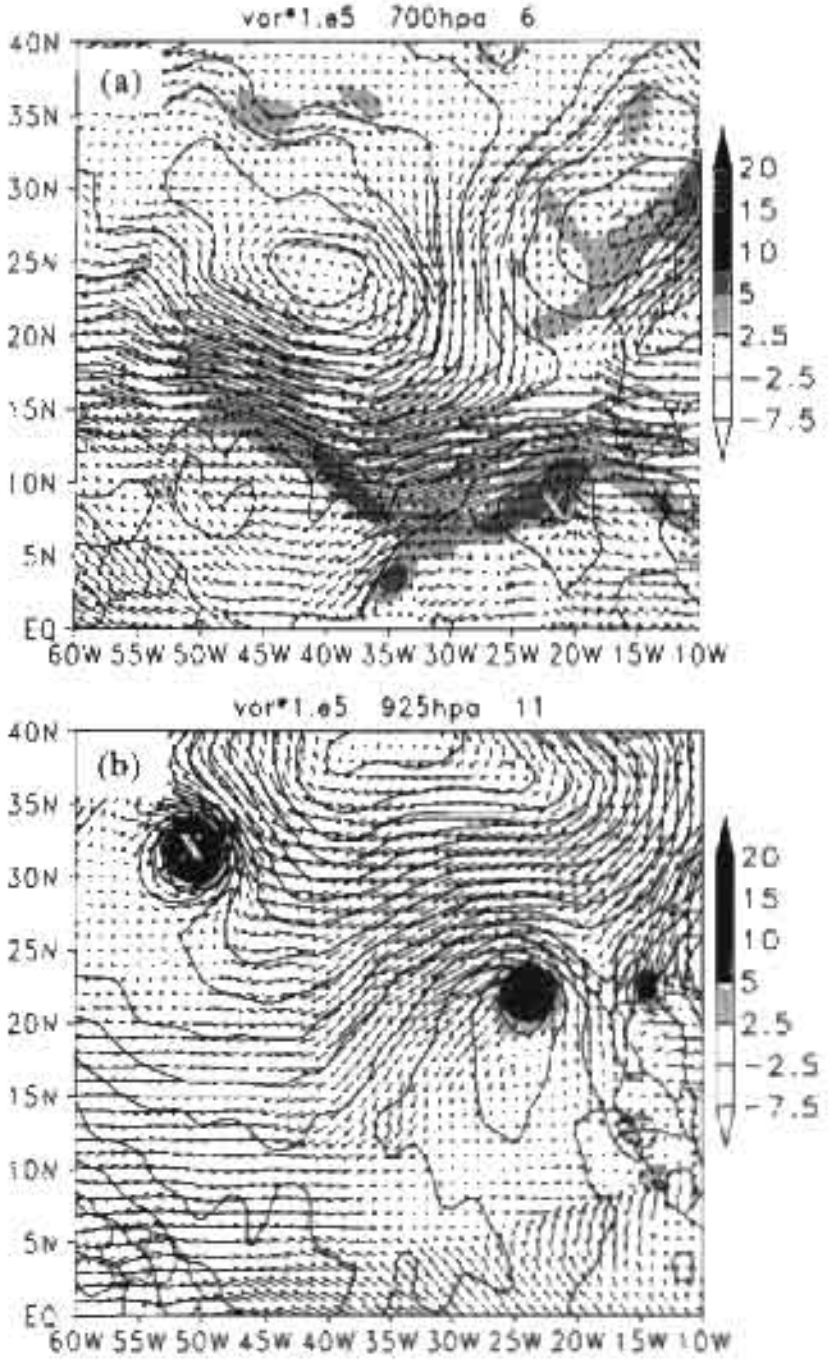


Figure 4: The usefulness of different variables in tracking systems in different situations (a) Vorticity is useful in cases with broad waves (b) For mature cyclones, the central closed height line provides a good estimate of where the center is located (c) Systems with broad vorticity and height signatures can be tracked by the approximate center of circulation denoted by the wind vectors (d) Sometimes the center of the selected system is ambiguous and no variable is most useful

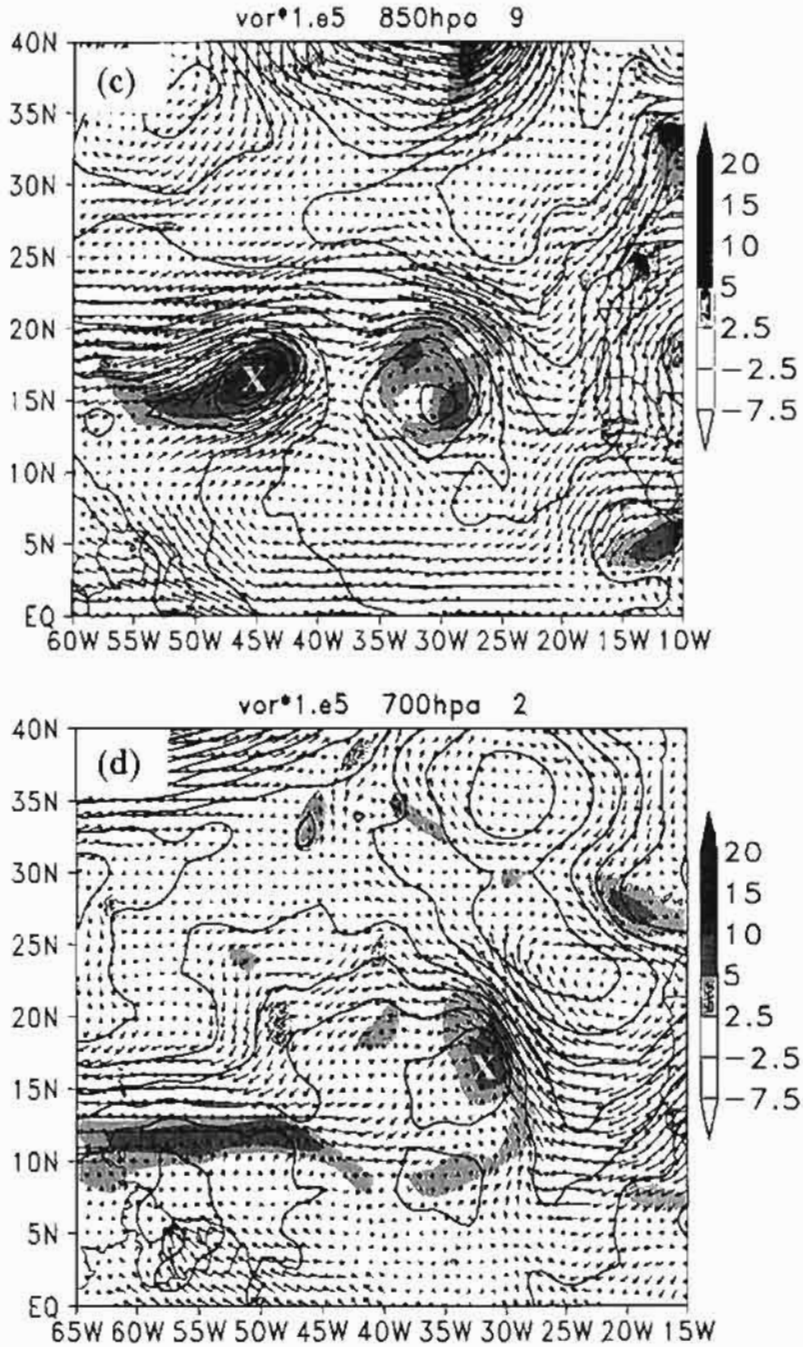


Figure 4 continued.

From the number of cases with various forecasting leading times, it was concluded that all three variables were important to best find the center of the system. Thus, all three variables were necessary for accurate cyclone tracking.

2.3.4 Tracking levels

Similarly, there is no standard height level used to track cyclones in previous studies. In order to examine which level is most effective for tracking, the cyclone tracking experiments were conducted at different pressure levels. Although it is believed that the upper level atmospheric conditions are important for tropical cyclone genesis, the primary focus of this study is to determine if the model predicts genesis, and therefore only lower pressure levels are considered (e.g., 925, 850, and 700 hPa).

Cyclone tracking was performed at each of these pressure levels (925, 850, and 700 hPa) and compared to the best track data. For stronger systems, the tracks were virtually identical at the different pressure levels. In weaker systems, particularly before genesis, there were variations of the track amongst the different levels, but no systematic pattern was identified. Fig. 5 shows two opposing examples of tropical systems approximately 3 days before they were designated tropical depressions. Fig. 5a shows a very weak wave (pre-Gordon on 8 Sep 2006) that had widely varying signatures at different vertical levels. Meanwhile, the pregenesis stage of Helene (on 10 Sep 2006) was well organized, and the tracks were virtually the same at each pressure level (Fig. 5b).

In general, the system representation was stronger at lower levels, but in some cases weaker waves were better depicted at 700 hPa. Cyclones also were forecast to deteriorate from the top down, which suggested that the stronger cyclones tended to be

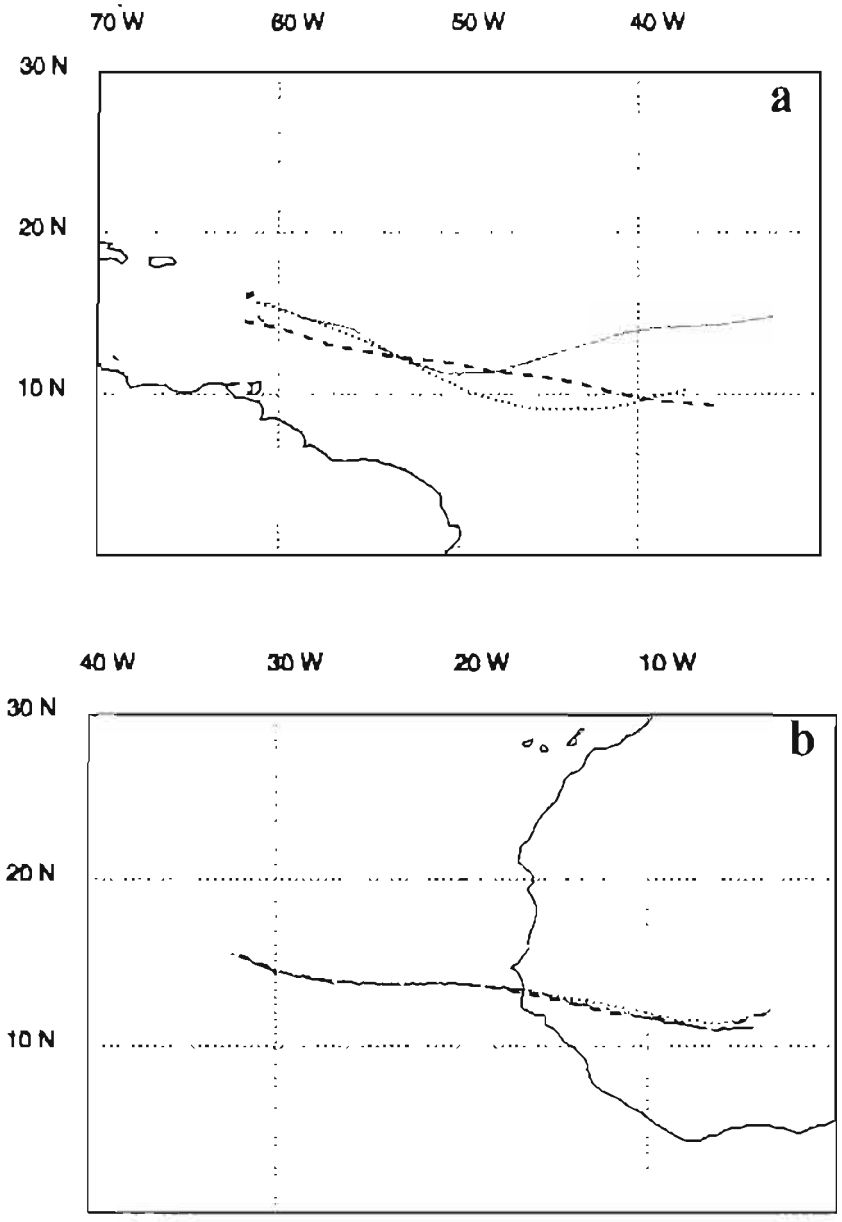


Figure 5: Example of variance in tracks at different atmospheric levels. (a) Ensemble mean 0-120 h forecast tracks from 66 h before Gordon was designated a tropical depression (0000 UTC 08 Sep 2006). (b) Ensemble mean tracks from 60 h before Helene was designated a tropical depression (0000 UTC 10 Sep 2006).

deeper. Similar discontinuities between pressure levels in weak tropical systems were also seen when comparing the manual tracks to the best track data.

Table 1 illustrates the conceptual magnitude of the track errors for all cases, with different forecast lead times at different pressure levels. After assessing and comparing the benefits of all levels, the tracks at 850 hPa were deemed to minimize the track errors and are used for all the cases in this thesis. In fact, the use of 850 hPa as the tracking level also agrees with common justification: due to effect from the ocean or land surface, diurnal vorticity maxima may be present at 925 hPa. In addition, weaker systems may not be depicted at 700 hPa since a weaker system is usually shallower in terms of its entire depth.

2.3.5 Daily genesis potential

Originally developed by McBride and Zehr (1981), and used recently by Kerns et al. (2008), the daily genesis potential (DGP) is a tool to assess and track developing tropical disturbances. By thermal wind balance, the difference in relative vorticity within a layer approximates the strength of the warm core (if one exists) and indicates the likelihood of tropical storm formation. Commonly, the DGP is defined by the raw difference of relative vorticity at 925 and 200 hPa: $DGP = \zeta_{925} - \zeta_{200}$. By this definition, a high DGP is consistent with a deep warm core, and indicates that genesis is proceeding.

The DGP was plotted for tracking in the same manner as vorticity to determine its usefulness as a tracking variable. Compared with low-level vorticity, the DGP produced comparable tracks to the other tracking variables. However, it also added track diversity when tracking weak tropical systems. Fig. 6 shows a typical track that followed this pattern. On the other hand, as a proxy for strength, the DGP provided inconsistent results.

Table 1: Comparison of track errors amongst the three pressure levels for each forecast lead time relative to a genesis time. Negative numbers denote the forecasts start number of day(s) before the genesis time and positive numbers represent the forecasts start number of day(s) after the genesis time. "x" denotes the track error less than 150 km. "v" represents the track error greater than 150 km but the error is 50 km (a half of model grid size) of the minimum track error among the three pressure levels.

Case	Debby					Ernesto						Florence						Gordon						Helene					
Forecast lead time	-3	-2	-1	0	1	-3	-2	-1	0	1	2	-3	-2	-1	0	1	2	-3	-2	-1	0	1	2	-3	-2	-1	0	1	2
925 mb			x	x	x		x	x			x	x	x		x	x	x	v					x	v	v	v	x	x	x
850 mb		x	x	x	x	v	x	x	x	x	x		x	v	x	x	x	v	v	v	x		x	v	v	v	x	x	x
700 mb	v		x	x	x	v		x	x	x	x		x		x	x	x			v	x	x	x	v	v		x	x	x

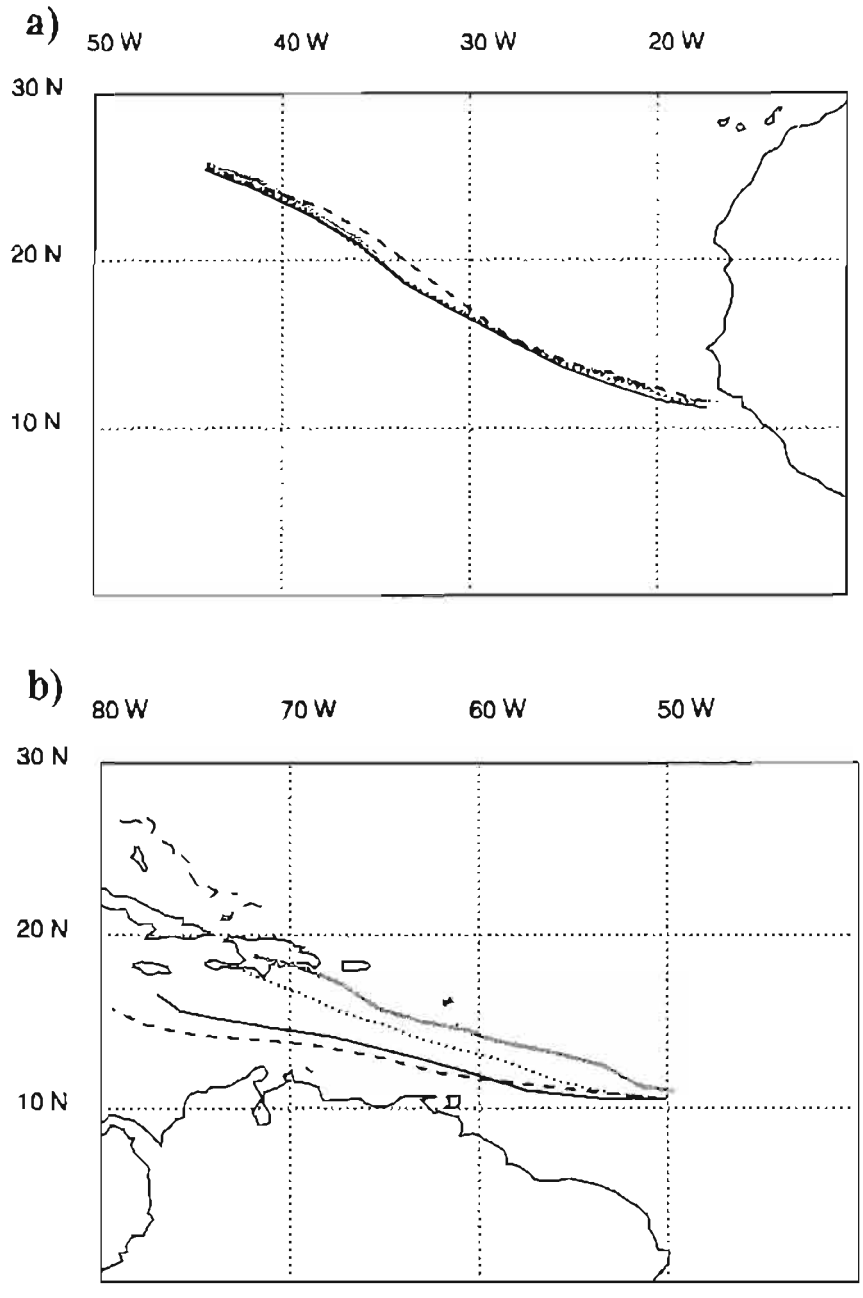


Figure 6: Ensemble mean tracks at different pressure levels (dashed curve for the track at 700 hPa, thin solid curve for 850 hPa, and dotted curve for 925mb), compared with the DGP-tracked ensemble mean (thick gray solid curve). (a) The 5-day forecast from 21 Aug for Debby shows the case of a stronger system. (b) The forecast from 23 Aug for Ernesto shows the case of a weaker system, and the DGP track deviates far northward of all of the individual levels.

Based on these case studies, a threshold value could not be established to justify genesis. Therefore, although the DGP is a useful parameter, it does not add significant value for our specific purpose in tracking cyclone genesis and is not used for cyclone tracking in this study.

2.4 Case studies

Based on the methods and criteria set up in Section 2.3, cyclone tracking was performed for five named cyclones and two nondeveloping waves. The systems were tracked in ensemble forecasts from three days in advance until 2 days after being designated a tropical depression by the NHC. The tracking results for all five named cyclones are summarized in Table 2. A detailed description for each case is given as follows. The tracks are plotted for each forecast, with the locations of predicted genesis denoted by an "X" on the track, which gives an indication of the intensity forecasts (e.g., the timing and number of members).

2.4.1 Debby

The African easterly wave (AEW) that developed into Debby was disorganized and somewhat ambiguous over land, and was thus not well resolved in the model. After moving offshore though, circulation developed quickly. Debby was classified as a tropical depression at 1800 UTC 21 Aug 2006. It evolved into a tropical storm at 0000 UTC 23 Aug 2006. Due to dry and stable air, along with marginal sea surface temperatures (SST), the cyclone never intensified into a hurricane. Shear associated with an approaching upper-level trough eventually caused the cyclone to dissipate (Franklin 2006).

Table 2: Predictability of each cyclone for different lead times relative to the system being designated a tropical depression by NHC. Values represent number of members (out of 14) predicting genesis (G), vortex-like development (V), premature dissipation (D), and nondevelopment (N), respectively. Tracking results from operational GFS forecast are also shown.

Case	**Debby**		**Ernesto**		**Florence**		**Gordon**		**Helene**	
Forecast Lead Time (days)	Number of ensemble members in G/V/D/N	GFS	Number of ensemble members in G/V/D/N	GFS	Number of ensemble members in G/V/D /N	GFS	Number of ensemble members in G/V/D/N	GFS	Number of ensemble members in G/V/D/N	GFS
-3	6/4/2/2	V	0/1/0/13	N	10/3/1/0	V	0/2/0/12	N	14/0/0/0	G
-2	4/7/2/1	G	0/3/0/11	G	7/3/0/4	V	0/0/0/12	N	11/2/0/0	G
-1	6/3/4/1	G	0/9/0/5	V	14/0/0/0	G	0/7/0/7	N	12/0/2/0	G
0	10/0/4/0	G	1/13/0/0	V	12/2/0/0	G	0/4/5/5	D	13/1/0/0	G
1	1/6/0/7	V	1/13/0/0	V	13/0/1/0	G	3/10/1/0	G	14/0/0/0	G
2	n/a	n/a	5/6/3/0	G	13/1/0/0	G	14/0/0/0	G	14/0/0/0	G

In the 5-day model forecast started from 0000 UTC 19 Aug (Fig. 7a), very few ensemble members accurately predicted the evolution of Debby. A majority of the members predicted a track that was well south of the best track. In addition, most of the members that predicted genesis failed to ultimately strengthen the cyclone. In the forecasts from 20 Aug 2006, all ensemble members agreed better about developing a cyclone (Fig. 7b). As the forecast time approached genesis, most members followed a similar track. Consistency between different ensemble members improved as the initialized tropical disturbance became better organized closer to genesis. Tracks were aligned and spread was minimized for the forecasts starting from 21 and 22 Aug (Fig. 7c and d). In particular, the forecast tracks from 22 Aug started from the same location due to the NCEP implementation of the vortex relocation after the system became a tropical depression (Fig. 7d). The ensemble mean was also very close to the best track on this day. For the forecasts started 0000 UTC 23 Aug 2006, most of ensemble members did not predict the northward turn of the cyclone at later forecast hours, but demonstrated an evenly spread ensemble field that generally encompassed the best track (Fig. 7e). Overall, the ensemble predictability of Debby's track showed tremendous improvement as the cyclone became more organized, especially after it developed into a tropical depression (Fig. 7a-e).

The GFS forecast from 0000 UTC 19 Aug did not show any development from the predominant vorticity maximum, but rather showed a brief rapid deepening of an unrelated vorticity maximum to the southwest of the actual system's location (Fig. 7a). For the remainder of the forecasts, the GFS tended to depict a strong storm. The forecasts started from 20 and 21 Aug were slower and more northerly than the ensemble mean

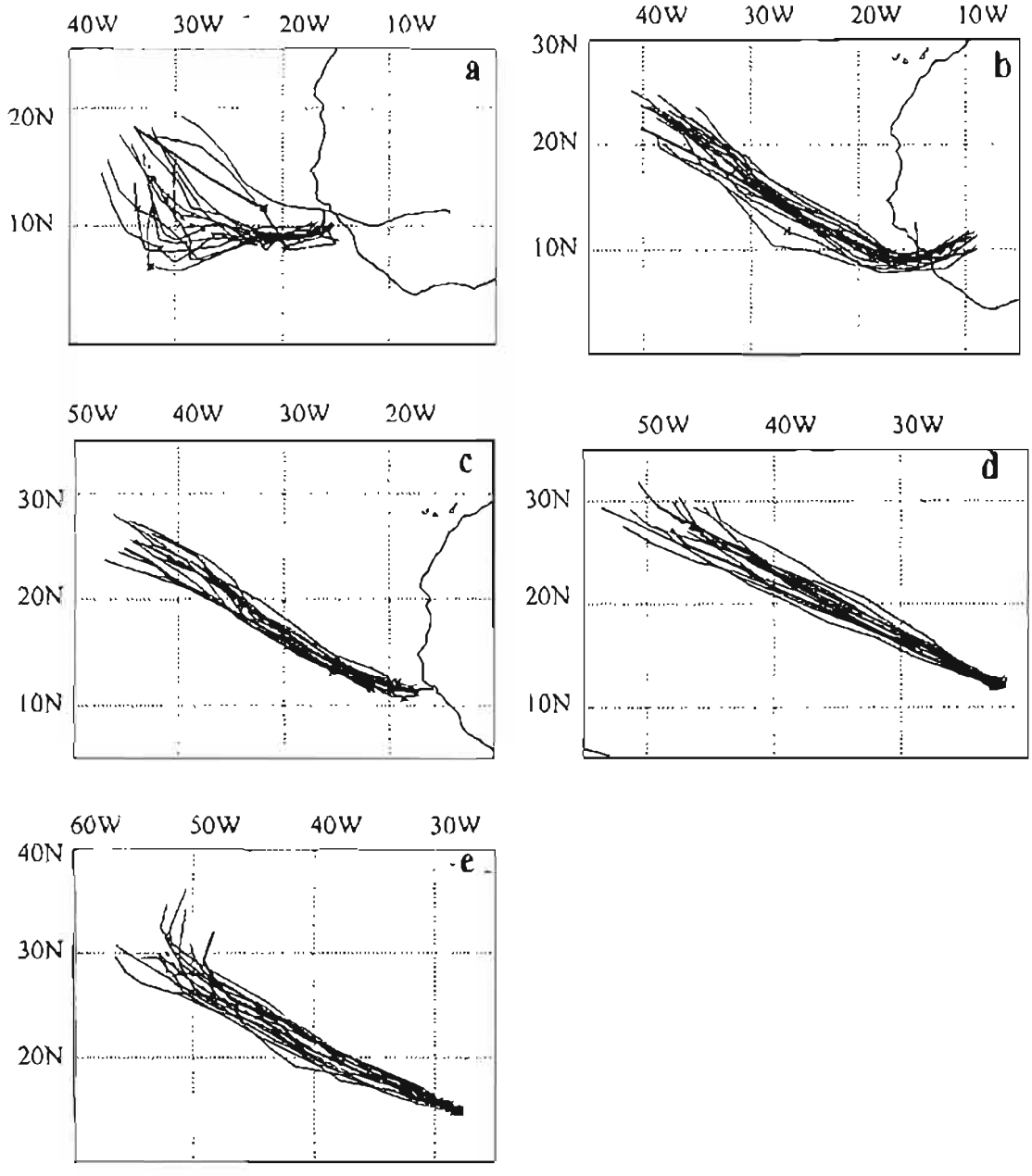


Figure 7: The tracks of 0-120 h ensemble forecasts for Tropical Storm Debby (purple thin lines) from 0000 UTC 19-23 Aug 2006 (corresponding to Fig. 7 a-e, respectively), compared with the corresponding operational GFS forecast (blue line) and NHC best track (thick black line). Red dashed lines denote the ensemble mean. An "X" designates the forecasted genesis location.

(Fig. 7b and c), while the forecasts from 22 and 23 Aug align more closely to the ensemble members (Fig. 7d and e).

2.4.2 Ernesto

Ernesto formed from a weak tropical wave traversing the south-central North Atlantic Ocean. Convection increased as the system moved around the southern extent of a mid-level ridge. Tropical depression status was reached at 1800 UTC 24 Aug 2006, followed by tropical storm status at 1200 UTC 25 Aug 2006. The cyclone briefly reached hurricane status on 27 Aug but quickly weakened as it interacted with terrain in Haiti. Strengthening was limited as the cyclone crossed Cuba and made landfall in Florida. Moving northeastward due to a weakness in the mid-level ridge, Ernesto strengthened once again as it moved out over the waters of the open Atlantic. The cyclone nearly reached hurricane force before making final landfall in North Carolina (Knabb and Mainelli 2006).

Surprisingly, all forecasts before the genesis time did not develop a cyclone at all, although the number of vortex-like systems increased with decreasing forecast lead time. From the 5-day forecasts started on 22 Aug, all the members forecasted the wave to remain weak. Ensemble forecasts also moved the wave too quickly to the west, with small ensemble spread until the later stages of the forecast (Fig. 8a). While more members produced weak vortex-like circulations in the forecasts that started on 23 and 24 Aug, predictions of a well-developed tropical storm remained nonexistent (Fig. 8b and c). The track of the wave and vorticity maximum remained much farther south than where the actual cyclone tracked and was much faster than the actual cyclone. In forecasts from 25 Aug, many of the ensemble members continued to predict a westward track, even after

the system developed into a tropical depression (Fig. 8d). The ensemble mean showed a track into the Gulf of Mexico, while the actual storm made landfall in southern Florida. This pattern continued in forecasts from 26 Aug and no members produced a well-developed storm, and many completely dissipated it. Those members that held the system together showed deepening over the Gulf Stream offshore from the eastern United States. None of the ensemble members in the forecasts from 27 Aug (the last day tracked, when the cyclone initialized near Hispaniola) predicted the eventual landfall in North Carolina (Fig. 8f). Overall, the ensemble forecasts of Ernesto largely failed since most of the pregenesis forecasts only produced vortex-like tropical waves. While after genesis the track forecasts showed some improvement (Figs. 8 d, e, and f), the intensity forecasts remained too weak.

The GFS forecast starting from 0000 UTC 22 Aug predicted a track on the north side of the ensemble envelope (Fig. 8a). In the forecasts initialized on 23 Aug, the GFS forecasted development, but it occurred from a vorticity maximum that split off from the wave and developed northeast of the ensemble mean and the actual track. The GFS forecasts from 24-26 Aug were generally south of the ensemble mean (Fig. 8 c-e). Forecasts from these 3 days maintained circulation, but the tracks were closer to the ensemble mean than the actual track. However, the GFS forecast starting on 27 Aug (Fig. 8f) produced a storm with appropriate intensity and predicted a nearly-perfect track (as opposed to the ensemble members), although it moved slightly slower than the actual cyclone.

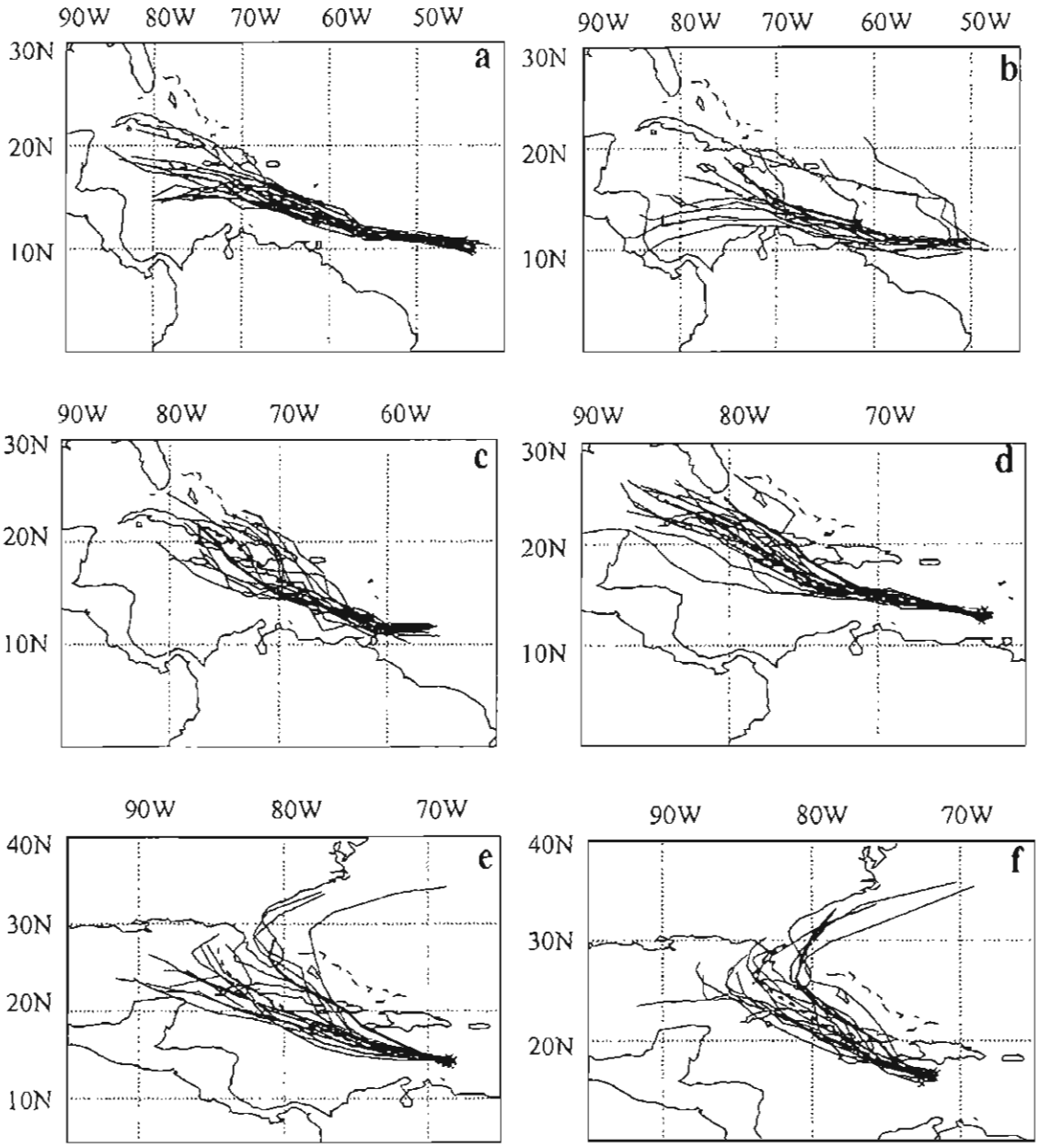


Figure 8: The tracks of 0-120 h ensemble forecasts for Hurricane Ernesto from 0000 UTC 22-27 Aug 2006 (corresponding to Fig. 8 a-f, respectively), compared with the corresponding operational GFS forecast and NHC best track.

2.4.3 Florence

Two tropical waves interacted to form Florence. The first wave moved slowly westward off the coast of Africa on 29 Aug, while the second followed at a much higher speed, overtaking the first. After convective organization, Florence was first classified as a tropical depression at 1800 UTC 03 Sep 2006. Initially very broad and containing multiple vorticity maxima under strong shear, the depression gradually strengthened to a tropical storm by 0600 UTC 05 Sep 2006. As shear weakened and the storm consolidated, hurricane status was reached on 10 Sep near Bermuda. The cyclone turned to the north, then northeast, and became extratropical as it moved into the waters of the North Atlantic (Beven 2006).

Florence presented a complex situation in the pregenesis ensemble forecasts. Specifically, there were three tropical waves over the central Atlantic, and the forecasts from different ensemble members handled them differently. In the ensemble forecasts, some members developed a cyclone from the first wave, while other members developed one from the second wave or from the merger of the first and second waves. On a few occasions, there would be a second merger with the third wave, or a cyclone would develop from the third wave itself. Some ensemble members accurately predicted genesis, but also generated an equal or even stronger storm to the northeast of Florence. In general, the continuity was inconsistent and the tracking was difficult in the pregenesis environment of these ensemble forecasts. These complexities were present in the forecasts from 1, 2, and 3 Sep, although the forecasts became more consistent as time progressed. For example, Fig. 9 a-c show a large ensemble spread present in the forecast from 1 to 3 Sep, with few members representing the best track. After the genesis stage

(forecasts from 4-6 Sep, Fig. 9 d-f), the cyclone tracks were well predicted by the ensemble forecasts, with smaller spread and a mean track close to the best track. Compared with Fig. 9 a-c, the ensemble forecast significantly improved after the wave strengthened to a tropical depression. The improved postgenesis track forecasts can be attributed to the NCEP vortex relocation scheme.

In the GFS control forecast, the third wave formed the largest cyclone in the forecasts starting from 1 and 2 Sep. For the forecast starting on 3 Sep, the GFS presented a complex and discontinuous solution that was similar to a subset of the ensemble members (Fig. 9 a-c). In the postgenesis phase (4-6 Sep), the GFS was more similar to the ensemble (Fig. 9 d-f).

2.4.4 Gordon

Gordon formed from a weak tropical wave in the central Atlantic. Wind shear associated with the upper-level trough from Florence limited development as it moved westward. As shear eased, the system reached tropical depression status at 1800 UTC 10 Sep and tropical storm status at 1200 UTC 11 Sep. It curved slowly northward within a weakness in the subtropical ridge. Shear diminished, and Gordon became a hurricane on 13 Sep, rapidly intensifying thereafter. After meandering in the central Atlantic, Gordon accelerated northeastward and eventually became an extratropical low on 21 Sep as it approached Portugal (Blake 2006).

For the 5-day forecast started from 8 and 9 Sep (Fig. 10 a and b), none of the ensemble members predicted the actual position of the system. Even when the ensemble showed a strengthening of the tropical disturbance, it was too far south or west of the actual track. Fig. 10a (forecasts from 8 Sep) shows how the majority of the tracks were

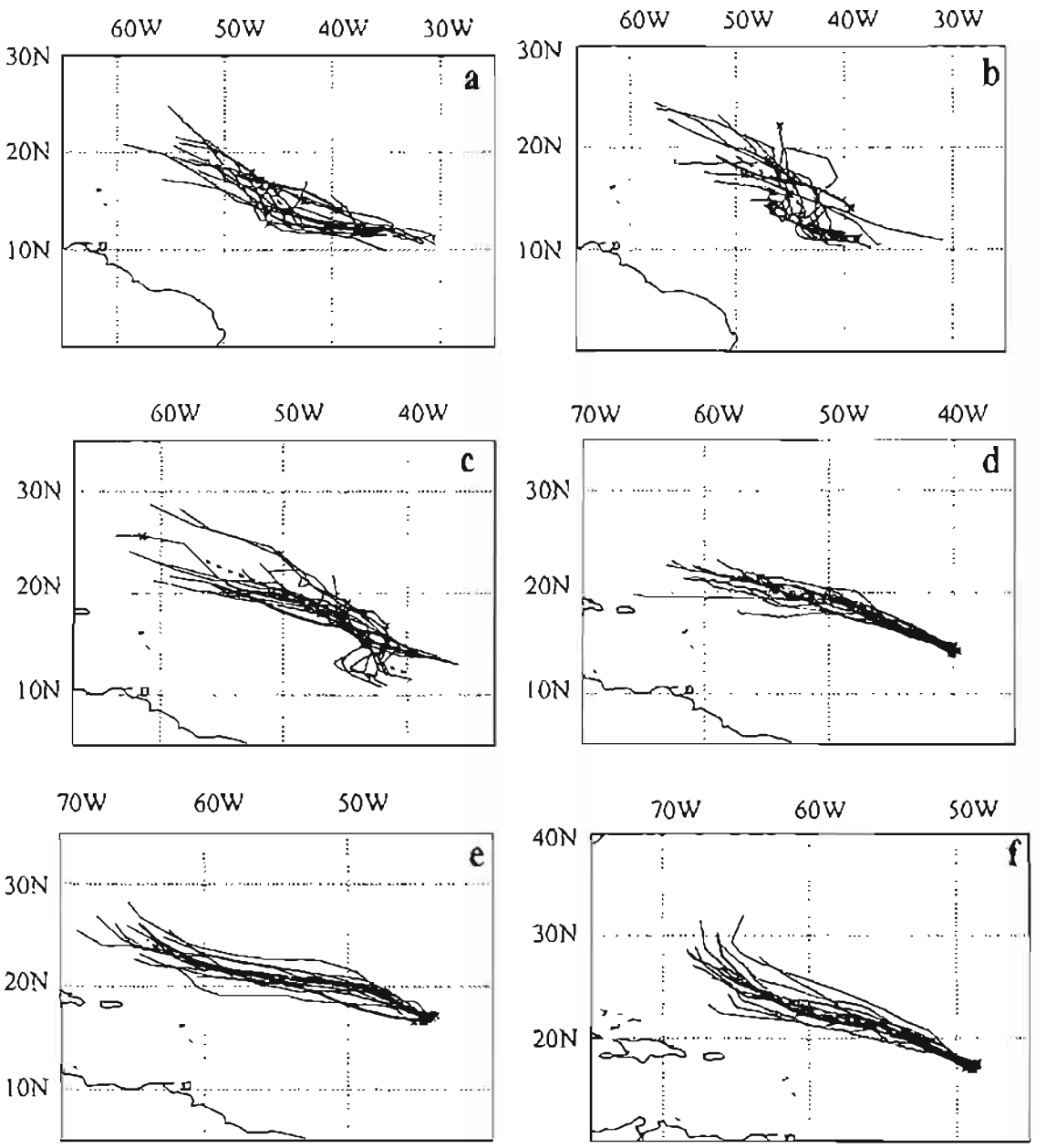


Figure 9: The tracks of 0-120 h ensemble forecasts for Hurricane Florence from 0000 UTC 1-6 Sep 2006 (corresponding to Fig. 9 a-f, respectively), compared with the corresponding operational GFS forecast and NHC best track.

well south of the actual genesis location. In most of these cases, the initial vorticity maximum drifted southward and merged into a strip of high vorticity. The forecast from 10 Sep (Fig. 10c) produced a wide array of tracks, with only one member correctly showing the system curving northward. By 11 Sep (after the tropical depression had developed), the ensemble forecasts correctly predicted movement (Fig. 10d). However, none of the ensemble members suggested that a strong hurricane would eventually develop. Also, ensemble forecasts were consistently too fast with movement, leading to erroneous forecasts of dissipation or absorption by mid-latitude troughs. Once again, the NCEP ensemble forecast showed significant improvement in track forecasts after the system became a tropical depression (Fig. 10 e and f).

The GFS forecasts from 9 and 10 Sep also produced a cyclone that was weak and too far south (Fig. 10 a and b). In the forecasts starting on 10 Sep, the original vorticity maximum dissipated, while another wave formed within the vorticity strip. From the forecast on 11 Sep, the GFS was somewhat similar to the ensemble mean (Fig. 10d). The GFS forecasts from 12 Sep proved to be very close to the actual track, while maintaining strength and circulation (Fig. 10e). Initializing on 13 Sep, the GFS track mirrored both the best track and ensemble mean, but was displaced to the west (Fig. 10f).

2.4.5 Helene

The Helene case was the easiest to track, as a strong AEW developed quickly as it moved westward off the coast on 11 Sep. It was classified as a tropical depression at 1200 UTC 12 Sep 2006. Although initially inhibited by easterly shear, the depression was upgraded to a tropical storm at 0000 UTC 14 Sep. Helene moved to the west-northwest, strengthened to a hurricane by 1200 UTC 16 Sep, and eventually became a category 3

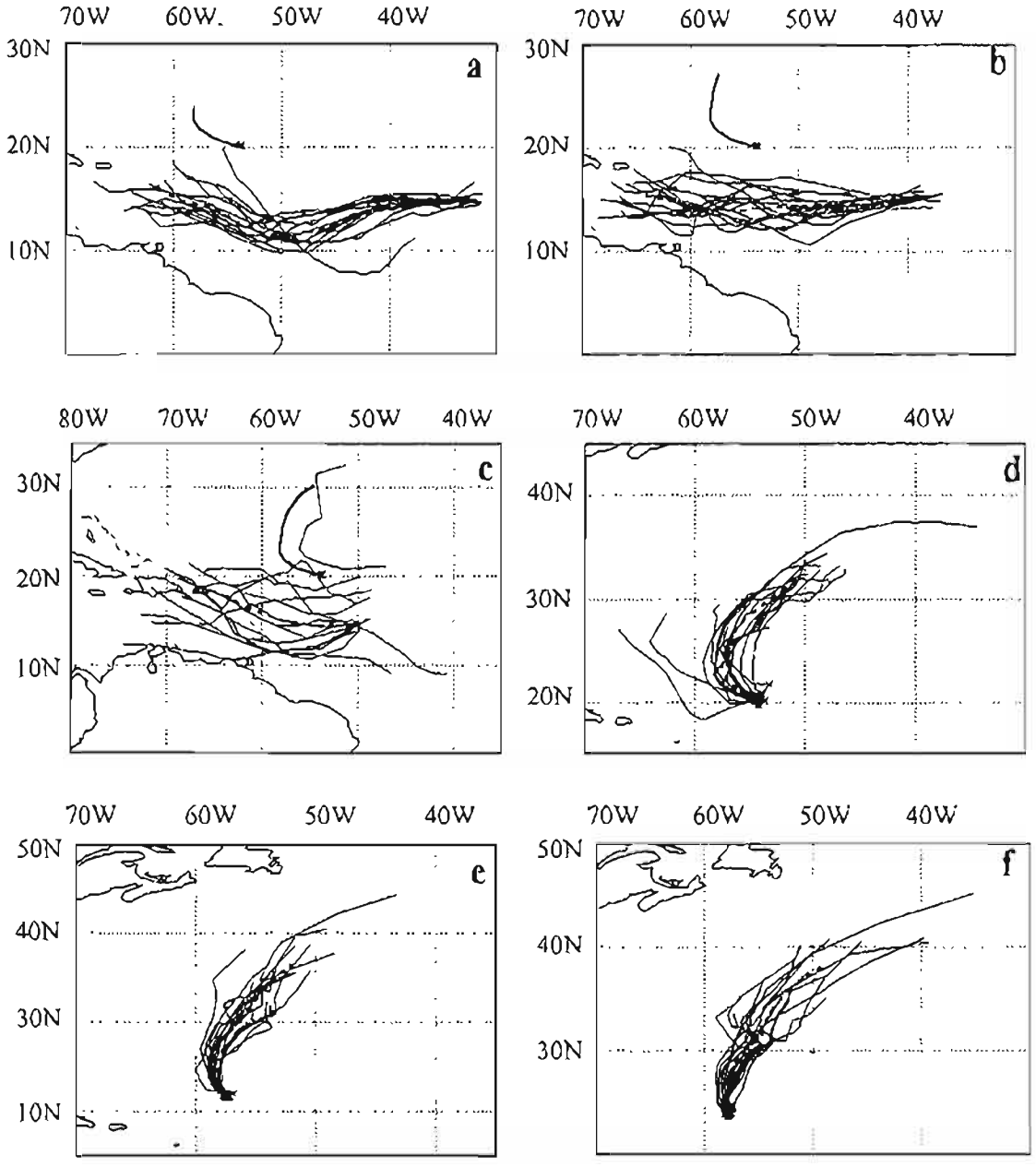


Figure 10: The tracks of 0-120 h ensemble forecasts for Hurricane Gordon from 0000 UTC 8-13 Sep 2006 (corresponding to Fig. 10 a-f, respectively), compared with the corresponding operational GFS forecast and NHC best track.

hurricane. Movement and speed were impacted by the location of Hurricane Gordon. An upper-level trough exiting the United States caused Helene to curve east of Bermuda. The cyclone became extratropical on 24 Sep (Brown 2006).

Even in the pregenesis phase over Africa, the vorticity maximum and wind circulation were well defined in the model initialization. The forecast from 0000 UTC 10 Sep showed a strong developing cyclone as soon as the AEW exited the coast, with all 14 members predicting genesis. In Fig. 11a, the track forecasts from 10 Sep are close to the actual track with little spread, although most members predicted slower movement than what actually occurred. The forecasts from 11 Sep (Fig. 11b) suggested a weaker cyclone, and some members dissipated the cyclone. Predicted cyclones in these forecasts generally moved slower and more northerly than the actual track with small track spread. From 12 Sep onward, the ensemble forecasts were generally accurate. The first forecast after tropical depression status was reached (0000 UTC 13 Sep) is shown in Fig. 11d, which shows the ensemble mean was close to the best track with a classical cone-like ensemble spread. Ensemble members handled the timing of the deepening with relative accuracy. However, the forecast from 15 Sep did not show the eventual western drift of the cyclone (Fig. 11f). Overall, all forecasts for Helene displayed a high degree of accuracy.

GFS tracks were similar to the ensemble mean for this particular case. One exception was the forecast initialized on 11 Sep (Fig. 11b), where the storm movement was much slower than the ensemble members and showed a northerly turn. Otherwise, the GFS forecasts were similar to the ensemble.

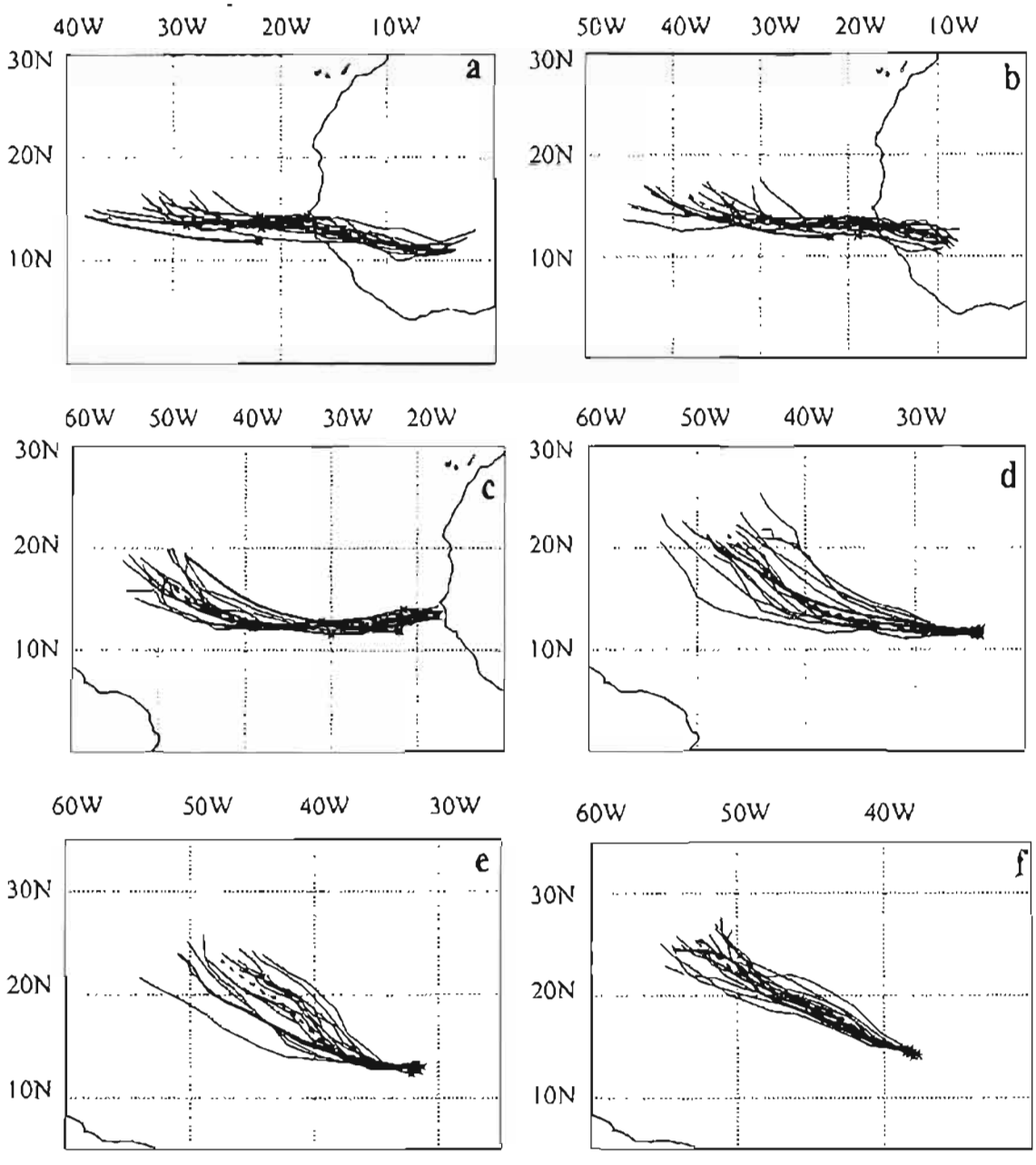


Figure 11: The tracks of 0-120 h ensemble forecasts for Hurricane Helene from 0000 UTC 10-15 Sep 2006 (corresponding to Fig. 11 a-f, respectively), compared with the corresponding operational GFS forecast and NHC best track.

2.4.6 "Wave 3"

Although the aforementioned five tropical storms all developed from AEWs, not all easterly waves develop into tropical storms. Some studies, such as Thorncroft and Hodges (2001), have sought to determine a relationship between AEWs and tropical cyclone genesis. While they found a correlation between the number of these mainly meridionally-oriented tropical waves and the subsequent number of tropical cyclones, not all the waves strengthen into cyclones. When an AEW (or commonly named "tropical disturbance") does not reach tropical depression status as determined by the NHC, it is considered a nondeveloping system. Forecasts that predict tropical cyclone genesis within these nondeveloping tropical waves are referred to as "false alarms." There were two nondeveloping cases during NAMMA. To evaluate whether the ensemble produced such false alarms, the two nondeveloping waves during this period are examined.

"Wave 3" was an AEW that was well-defined as it moved westward off the African coast. Tracking was performed between 23 and 27 Aug during which the wave travels from east to west. Afterward, it slowly weakened, never developing into a tropical storm. The ensemble forecasts handled the tracks of Wave 3 relatively well (Fig. 12). Tracks were tightly clustered, and closely followed the objective analysis of Zawislak and Zipser (2009, in press), except for the forecasts from 25 Aug, which took a more southerly track. Wave 3 had complex movement off the coast of Africa, with an interaction between one vorticity maximum moving westward, and another traversing southward down the coast from the north (seen more so at 700 hPa). With the weak structure, the mean tracks at various levels were diffuse, especially in forecasts initialized 24 and 25 Aug (850 and 700 hPa tracks diverge). Intensity was initially moderate. A

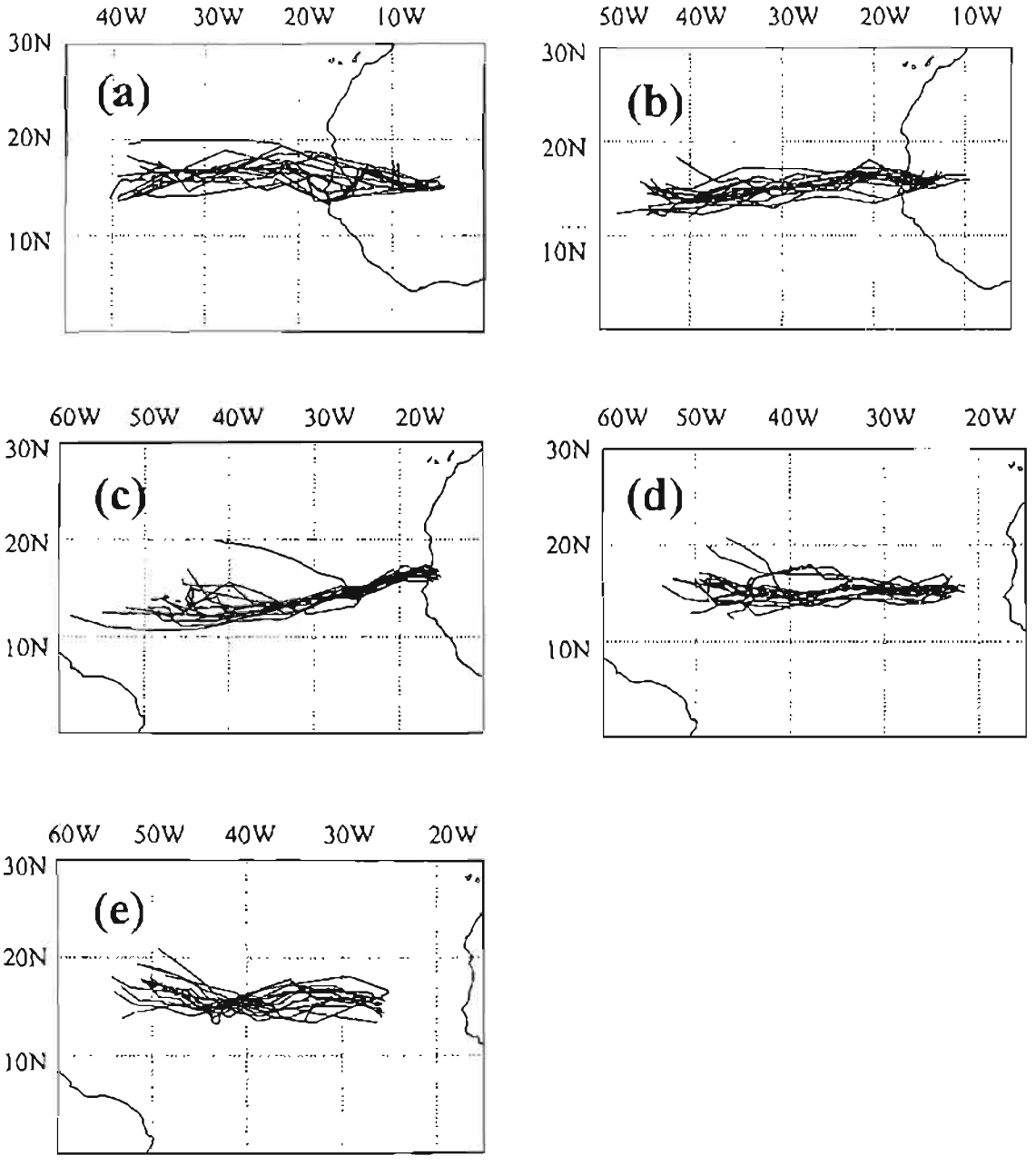


Figure 12: The tracks of 0-120 h ensemble forecasts for "Wave 3" from 0000 UTC 23-27 Aug 2006 (corresponding to Fig. 12 a-e, respectively), compared with the corresponding operational GFS forecast and NHC best track.

broad, weak circulation was evident initially in most of the forecasts, which by our criteria, constituted vortex-like structure. However, there were no forecasts of genesis or maintained strength (Table 2), meaning model performance for intensity was reasonable. While the operational GFS occasionally spun up small vorticity centers within the wave, none of them met genesis criteria (except one vortex-like case on 26 Aug), and most quickly dissipated.

2.4.7 "Wave 6"

"Wave 6" was another AEW that tracked due west from Africa. Its strength was much weaker than Wave 3 but was better defined at 700 hPa. Wave 6 could only be consistently tracked in the five-day forecasts from 7 through 10 Sep. Over time, this wave was forecast to dissipate and merge with a southern strip of high vorticity. The track forecasts had more spread than Wave 3, although movement was generally to the west with reasonable speed (Fig. 13). Wave 6 was weak and tracked differently at different atmospheric levels, especially from the forecast that started on 10 Sep (not shown). With weak initial intensity and a trend toward dissipation, none of the forecasts sampled predicted that the wave would strengthen significantly. In fact, even the few cases of vortex-like structure were very brief in nature. Tables 3 and 4 show the high rate of nondevelopment forecasts, indicating the accuracy of the ensemble in predicting these nondeveloping systems. The GFS also predicted a weak wave in all of the evaluated forecasts.

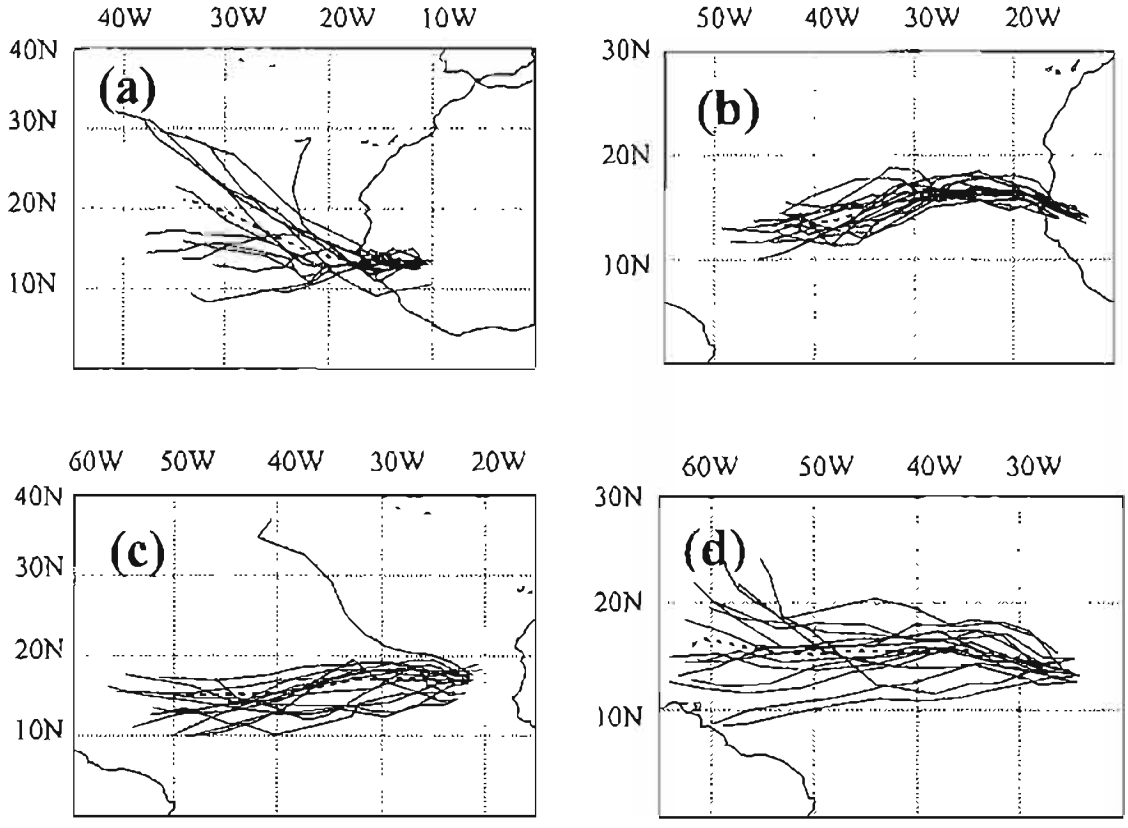


Figure 13: The tracks of 0-120 h ensemble forecasts for "Wave 6" from 0000 UTC 7-10 Sep 2006 (corresponding to Fig. 13 a-d, respectively), compared with the corresponding operational GFS forecast and NHC best track.

Table 3: Predictability of "Wave 3" in 5 day forecasts start from the different lead times. Values represent number of members (out of 14) predicting genesis, vortex-like development, premature dissipation, and nondevelopment, respectively.

Forecast Date	23 Aug	24 Aug	25 Aug	26 Aug	27 Aug
Genesis	0	0	0	0	0
Vortex-Like	10	8	14	10	8
Nondevelopment	3	6	0	3	6
Dissipation	1	0	0	1	0

Table 4: Predictability of "Wave 6" in 5 day forecasts start from the different lead times. Values represent number of members (out of 14) predicting genesis, vortex-like development, premature dissipation, and nondevelopment, respectively.

Forecast Date	7 Sep	8 Sep	9 Sep	10 Sep
Genesis	0	0	0	0
Vortex-Like	5	0	1	1
Nondevelopment	9	14	13	13
Dissipation	0	0	0	0

2.5 Overall evaluation

2.5.1 Predictability: Pregenesis vs. postgenesis

Table 2 shows the fraction of tropical cyclone genesis in ensemble forecasts started before and after genesis. The predictability of tropical cyclone genesis varies case by case as described in Section 2.4. For Debby, the number of members predicting genesis was fairly constant for all of the pregenesis forecasts. Locations of forecast genesis varied, with only the forecast initialized 19 Aug producing genesis forecasts that significantly deviated from observed. For Florence, at least 50% of members predicted genesis each day, with the smallest number of genesis predictions occurring with a two-day lead time. Most of the forecasts for genesis (from all lead times) took longer to develop than the actual cyclone (Fig. 9). Helene had the highest predictability, with at least 11 members predicting genesis each forecast cycle. For the forecasts initialized 10 Sep, most of the ensemble members developed the cyclone too quickly. Predicted genesis locations were spread evenly along the track in the forecasts initialized on 11 Sep, while more members took longer to develop a cyclone in the forecasts initialized on 12 Sep. As previously mentioned, none of the NCEP ensemble members met genesis criteria in pregenesis forecasts for Ernesto and Gordon (the GFS only predicted genesis on one day for Ernesto).

In most cases, the postgenesis (0 to 2 days from tropical depression designation) intensity forecasts were adequate and an improvement over pregenesis forecasts, but the ensemble system performed better for some cases than others. For Debby, all ensemble members met genesis criteria (although four predicted premature dissipation) in forecasts initialized on the day of genesis (22 Sep). The number of members predicting genesis

significantly dropped off the next day, with initial intensity estimates much weaker. The ensemble continued to depict weak intensity for Ernesto in the postgenesis forecasts. Only one member met genesis criteria on 25 and 26 Sep, and the genesis predictability only improved to five members on 27 Sep. Almost all members met genesis criteria in the postgenesis forecasts for Florence. The forecasts for Gordon continued to initialize with weak intensity for forecasts from both 11 and 12 Sep. However, all members showed accurate structure in forecasts from 13 Sep. As with Helene's pregenesis forecasts, the postgenesis forecasts also predicted a strong cyclone. The higher resolution GFS forecasts performed better in the postgenesis forecasts. Model initialization met genesis or vortex-like criteria for all cases and forecasts (with one case of premature dissipation).

Table 5 shows the overall probability of genesis in the ensemble forecasts over all five cases. Combining predictions of genesis and vortex-like cases, the probability for development is over 50%. The probability of the nondevelopment is less than 45% for the five named cyclones. Combining all three lead times to produce an overall average, the ensemble forecast had a genesis/vortex-like predictability of 73%, with the overall rate of 61 % for the forecast in pregenesis phase (-3 to -1 day lead time) and 87% for the forecast in the postgenesis (0 to 2 day lead time). These results show that the ensemble forecast offered a reasonable indication of the possibility of tropical cyclone genesis in these cases.

The track forecast was significantly improved in the postgenesis phase for all the cases (Fig. 7-11), largely because the implementation of the NCEP vortex relocation scheme in the operational ensemble forecasting system. In addition, both GFS and ensemble forecasts handled the two nondeveloped tropical waves very well. None of the

Table 5: Probability of genesis in ensemble for each lead time (forecast from number of days), combined for all five named cyclones.

Lead Time (days)	Genesis (G)	Vortex Like (V)	**G+V**	Dissipation (D)	Nondevelopment (N)	**(N+D)**
-3	30/70	10/70	**40/70**	3/70	27/70	**30/70**
-2	22/70	15/70	**37/70**	2/70	31/70	**33/70**
-1	32/70	19/70	**51/70**	6/70	13/70	**19/70**
0	36/70	20/70	**56/70**	9/70	6/70	**15/70**
1	32/70	29/70	**61/70**	2/70	7/70	**9/70**
2	46/56	7/56	**53/56**	3/56	0/56	**3/56**

forecasts predicted a false alarm of tropical cyclone genesis in any of the different forecast lead times (Tables 3 and 4).

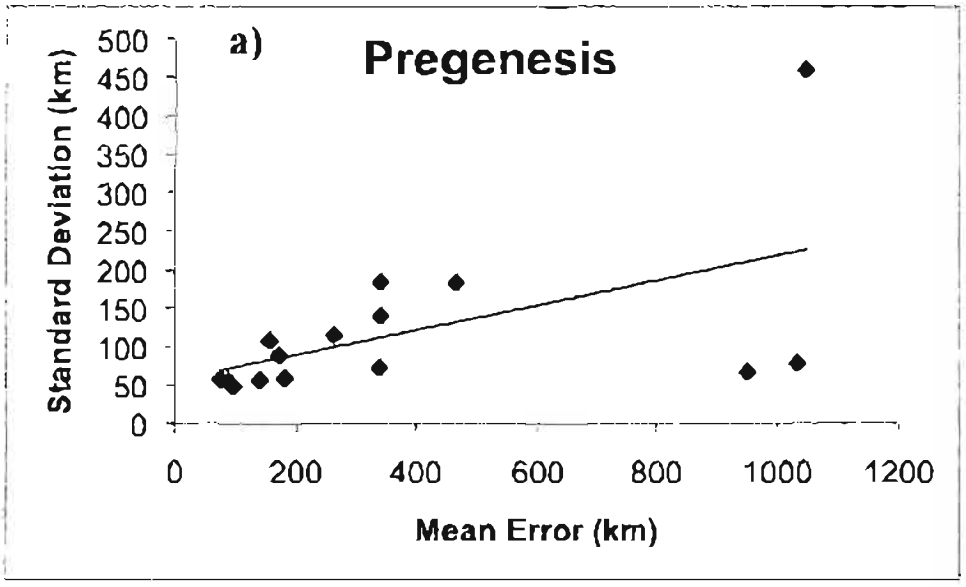
2.5.2 Accuracy of ensemble track forecast vs. ensemble spread

Goerss (2000) states that the ensemble spread could be a good indication of the predictability of the ensemble track forecast. Specifically, the ensemble spread is an approximation of the upper bound of the forecast track error. In order to confirm this useful point with the studies in this paper, the ensemble mean track errors are compared with the standard deviations of the ensemble members relative to the ensemble mean. The values of track errors show a decreasing trend of the errors with the reduction of lead time of the forecasts in majority of the cases, with the exception of Florence, in which case the track errors remained fairly constant. Track standard deviations are also varied with time, with Ernesto producing the highest values. For the forecasts that started in the pregenesis phase, Gordon had higher ensemble mean track error but lower standard deviation of the ensemble spread, proving that all ensemble members were consistently inaccurate. Table 6 summarizes the mean and standard deviation findings for each case and lead time relative to genesis.

To examine the relationships between the mean track error and ensemble spread, a linear fit has been applied to the two variables, and the coefficient of determination (R^2, or the square of the correlation) has been calculated to show how much of the variation in mean error is explained by the spread. Fig. 14 shows the correlations between the mean track error and standard deviation of the spread among all cyclones (averaged for each forecast cycle). The forecasts were divided into pre and postgenesis subsets. In the pregenesis subset, there is only minimal correlation overall, with an R^2 value of 0.29.

Table 6: Ensemble mean error and spread (km) for all cyclones. Lead time denotes the forecast from the number of days before the system was designated a tropical depression.

Lead Time	Debby		Ernesto		Florence		Gordon		Helene	
(days)	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean
-3	89.37	371.1	226.3	431.6	135	213.3	67.61	943.94	177.6	334.6
-2	58.31	127.5	123.9	128.9	80.9	107.8	311	1025.1	139.5	339.2
-1	51.91	87.1	64.53	118.8	54.62	200	411.8	1005.8	107.5	267.7
0	60.78	109.4	149.3	163.7	58.62	93.39	82.88	127.81	37.58	84.15
1	104.5	139.1	162.6	188.5	60.16	108.7	158.3	207.51	150.5	161.8
2	n/a	n/a	145.1	198.5	96.67	109.2	141.1	150.03	56.55	66.78



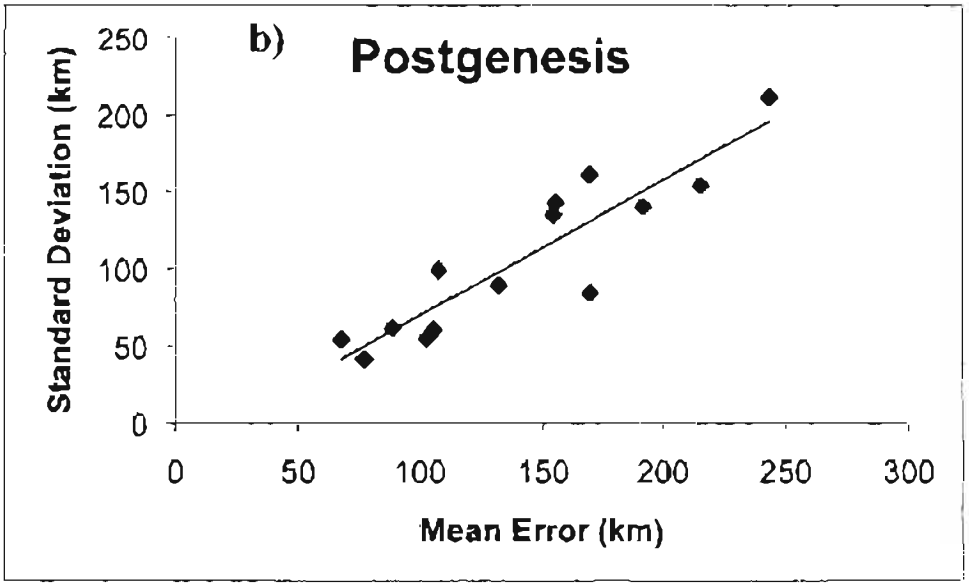


Figure 14: Scatter plots of ensemble mean error (km) vs. ensemble spread (represented by standard deviations of the ensemble members relative to the ensemble mean in km) averaged for each forecast period (1 to 5 days) for all cyclones. Forecasts are divided into two groups by the forecast lead time in (a) pregenesis, and (b) postgenesis.

However, by removing the outlier values (high mean error, low spread points on the right-most side of Fig. 14a) that resulted from the pregenesis forecasts of Gordon, the R^2 increases to 0.69, and the correlation between the mean track error and standard deviation becomes more obvious. The regression for the postgenesis forecasts is much more impressive, with 83% of the variance explained. These results from Fig. 14 confirm Goerss's (2000) conclusion that the ensemble spread is an indication of the uncertainty in the mean track.

2.5.3 GFS vs. Ensemble

The performance of the higher resolution GFS control forecast also varies between cases. For all five cases, a total of 29 forecasts have been evaluated. In 14 of the forecasts, the GFS track forecasts were very close to the track of the ensemble mean. For the forecasts in which the ensemble mean and GFS differed, the GFS resembled the best track data on two cases, while the ensemble mean was close to the best track six times. This difference may indicate additional value in the ensemble for tracking purposes. For another eight instances, GFS forecasts did not compare closely to the ensemble mean or actual track. In five forecasts, the GFS forecasted a track that significantly deviated from any ensemble member (the track was not within the ensemble "envelope").

Further investigation found that including the GFS as an additional ensemble member usually did not change the overall ensemble mean significantly, even when the GFS had a large deviation from the main ensemble members. The small difference results from the ensemble size of 14, which is much larger than the sample of a single control forecast. However, even with a single forecast, the overall performance of GFS in predicting tropical cyclone genesis is equal or better than the ensemble forecasts. Using

the overall genesis plus vortex-like prediction rate as a metric for pregenesis intensity forecasts, the GFS predicted tropical cyclone development 73% of the time, which is compatible to the overall ensemble accuracy of 73% for all cases and 61% for the ensemble forecasts in the pregenesis phase. The advantage of the GFS model forecast could be attributed to the fact that the forecasts have been produced at higher resolution. This advantage can also be seen in the forecasts from postgenesis phase. There were only three cases of vortex-like structure found in GFS: the Debby forecast from 30 h after genesis and Ernesto forecasts from 6 and 30 h after genesis. Additionally, the forecast from 6 h after the genesis of Gordon showed premature dissipation. In all other cases, the GFS correctly showed the storm structure according to the genesis criteria in this paper. Except in two examples, the GFS fell into the same intensity forecast category (genesis, vortex-like, etc.) as the majority of the ensemble members, which shows that the ensemble produces reasonable intensity forecasts on the whole.

2.5.4 Comments on storm strength

Tropical cyclone intensity, both pre- and postgenesis, was usually a good indicator of forecast accuracy. Helene was one of the strongest and most well-developed cyclones and was forecast the best by the ensemble. Florence was also well forecast by the ensemble after it organized. Debby was forecast poorly when it was a weak wave, but as Debby developed into a tropical storm, the forecast accuracy improved in the ensemble forecasts. Ernesto and Gordon were weak storms, at least initially, and were forecast poorly. Another factor for these two cases may have been their size. Even after development, their sizes were both small compared to other cases, so the ensemble forecasts may have had trouble resolving them. Gordon became a strong hurricane later

in its life cycle, but none of the ensemble members predicted it. It should be noted that the two poorly forecasted cyclones (Ernesto and Gordon) formed in the western North Atlantic, while the other three cyclones formed directly from AEWs. The differences in the pregenesis environment may also play an important role in forecast accuracy.

2.6 Summary

NCEP global ensemble forecasts for five developed and two nondeveloped tropical systems from the 2006 North Atlantic hurricane season were evaluated. The primary focus of the study was to determine how skillfully the ensemble performed in predicting the genesis and evolution of the tropical systems. Various track methods from previous studies were compared and evaluated based on their effectiveness in tracking tropical cyclone development. For this study, each case was tracked using a manual method that utilized vorticity, wind, and geopotential height at 850 hPa.

Overall, the ensemble forecasts have high probabilities of genesis for the three strong cyclones (Debby's mature phase, Florence, and Helene), but failed to accurately predict the pregenesis phase of the two weaker cyclones that formed farther west in the Atlantic Ocean (Ernesto and Gordon. Statistically, the overall rate of predictability for the genesis forecasts in the fived named cyclones was slightly above 70%. The forecast uncertainty generally decreased with the reduction of the forecast lead time, since the short-range ensemble forecasts tend to be more accurate. In contrast, the skill of ensemble track forecasts was significantly improved for matured cyclones, mainly because of the implementation of the NCEP vortex relocation scheme, which makes an accurate initial cyclone location for all ensemble members.

Although the overall performance of the high resolution GFS forecasts was equal or better than the ensemble forecasts in almost half of the cases of tropical cyclone genesis, the ensemble forecasts added value in predicting the probability of the tropical cyclone genesis. In addition, the ensemble spread could be a good indication of the forecasted track errors. Probability of detection for nondeveloping waves was high in both GFS and ensemble for two cases presented in this chapter.

CHAPTER 3

NOGAPS ENSEMBLE PREDICTION OF TROPICAL CYCLONE DEVELOPMENT: EVALUATION DURING T-PARC/TCS-08

3.1 Introduction

During the months of Aug and Sep 2008, a multinational field campaign commenced in the Western Pacific tropical basin. Under the umbrella of The Observing-System Research and Predictability Experiment (THORPEX) and the THORPEX Pacific Asian Regional Campaign (T-PARC), the Tropical Cyclone Structure Program (TCS-08) had investigated the mechanisms and predictability of tropical cyclone formation and development. The observation period of T-PARC consisted of targeted data gathering to improve initial conditions and model forecasts of tropical cyclone formation, track, intensity, structure change, and extratropical transition. Since this field experiment occurred in a different basin (Western Pacific vs. Atlantic Basin), it offered an opportunity to examine how ensemble systems would perform with different ambient conditions. The Naval Research Laboratory (NRL) was an integral part the campaign, and their global modeling system, the U.S. Navy Operational Global Atmospheric Prediction System (NOGAPS) model, was one of the key model systems used for field campaign support. The NOGAPS ensemble system ran independently with two configurations: multiple initial conditions and stochastic convection (described later). It

also played a role during the field program to guide decisions for targeted observations. In this chapter, similar evaluations to those in Chapter 2 will be presented for the NOGAPS ensembles. Most importantly, the results will be compared from two different configurations of the ensemble: with and without stochastic convection.

Four typhoons that formed during T-PARC/TCS-08 comprise the primary cases studies in this chapter. In addition, there were several tropical storms, and one was chosen for evaluation (to correspond to Tropical Storm Debby from the Atlantic cases). Many nondeveloping tropical waves were observed in the basin during the field experiment. Two of these waves were chosen to be studied: one that was consistently weak and did not develop in numerical forecasts and one that was slightly stronger and was considered by the forecasters to be a candidate for development due to numerical forecast for organization.

To maintain consistency, the naming system, tracks, and warnings follow those of the Joint Typhoon Warning Center (JTWC). Therefore, the named cyclones studied are referred to as Nuri, Sinlaku, Hagpuit, Jangmi, and Higos (in sequential order). The naming convention used for the TCS-08 field experiment will be used for the two nondeveloping cases: they will be referred to as TCS017 and TCS030 (which simply means they were the seventeenth and thirtieth convective systems designated during the experiment). While the official best tracks were unavailable at the time of this research, comparisons were made to the coordinates issued with the JTWC warnings (retrieved from http://www.typhoon2000.ph). Fig. 15 shows the JTWC tracks for the five named cyclones. Comparisons to actual positions will not be made for TCS017 and TCS030 since their locations were subjective and somewhat ambiguous due to their weak nature.

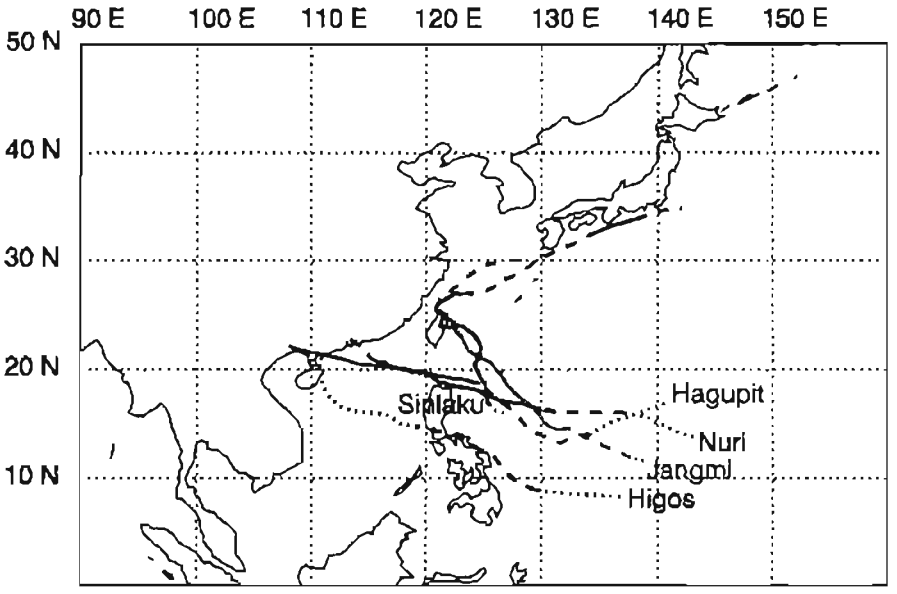


Figure 15: Joint Typhoon Warning Center (JTWC) best track for the five named cyclones investigated. Dotted lines represent tropical depressions, dashed lines represents tropical storms, and solid lines represent typhoons.

The work in this chapter also expands upon a recent study by Goerss and Reynolds (2008). They have evaluated the two configurations of the NOGAPS ensemble system (with and without stochastic convection) for tropical cyclone cases in the Atlantic Basin from 4 Jul to 31 Oct 2005. However, they only used nine out of the total 32 ensemble members, since previous studies showed further addition of members did not improve ensemble mean errors. They found that the addition of stochastic convection increased the independence of the forecast errors and thus resulted in a reduction of the ensemble mean error. In addition, the NOGAPS ensemble was compared directly to the GFS (NCEP) ensemble as well as several multimodel (consensus) ensemble forecasts. Among the cases studied, the GFS ensemble had smaller track errors in the short term (up to 48 h), whereas the NOGAPS had smaller track errors in longer range forecasts. While the single model ensembles were available to forecasters sooner than the multimodel ensembles, the relationship between the ensemble spread and ensemble mean track error was much weaker for the single model ensembles. A new consensus track tool that combines the single model ensemble means with the existing multimodel ensemble produced superior ensemble mean forecasts compared to any the individual members.

While the Goerss and Reynolds (2008) evaluation for NOGAPS ensemble mainly compared the ensemble skill with other ensemble systems, here we extend the evaluation to consider the ensemble skill during the pregenesis phase of the tropical storms, with the evaluation performed over all of the 32 members in two configurations of the ensemble (with and without stochastic convection). The main objective is to examine the relative accuracy of ensemble forecasts from these two different configurations in terms of their performance in the prediction of tropical cyclone genesis and development.

The evaluation results of the Western Pacific Basin cases in 2008 are discussed in the following sections. Section 3.2 describes the ensemble system. Tracking procedures are reviewed in Section 3.3. The tropical case studies examined are discussed in Section 3.4. An overall evaluation, including forecast accuracy, differences in the stochastic ensemble, and relationships between the ensemble mean and spread, follow in Section 3.5. Finally, Section 3.6 summarizes the results of this chapter.

3.2 Description of NOGAPS ensemble

3.2.1 Description of NOGAPS model

The NOGAPS is a global spectral model (Hogan and Rosmond 1991) that is currently the U.S. Navy's global operational forecast system. During the experiment, NOGAPS deterministic forecasts ran at a T239 horizontal resolution with 30 sigma levels vertically (T239L30). The model utilizes the Naval Research Laboratory Atmospheric Variation Data Assimilation System (NAVDAS), which is a three-dimensional variational system for data assimilation and model initialization. In order to make accurate tropical cyclone forecasts, the NAVDAS uses synthetic soundings, based on estimates of the cyclone's location and intensity for vortex initialization. These soundings are inserted at the center of the cyclone, as well as at locations two, four, and six degrees from the center. Winds are derived from a synthetic vortex with a radius of maximum winds of 50km. The NOGAPS is the basis of the NOGAPS ensemble systems that are evaluated in this chapter.

3.2.2 NOGAPS ensemble with perturbed initial conditions

From the deterministic NOGAPS model described above, the NRL developed an ensemble forecasting system with perturbed initial conditions. The most recent version of the NOGAPS ensemble system includes an ensemble transform method (referred to as ET, but not used here since "extratropical transition" also uses this acronym) to generate the initial perturbations, similar to that at NCEP (Bishop and Toth 1999). McLay et al. (2008) describe and evaluate the NOGAPS ensemble transform method and system. They found that this method was an improvement over bred vectors. The ensemble includes 32 members from the perturbed initial conditions and a control forecast from the original NAVDAS analysis, all with T119L30 resolution. Model error was not taken into consideration in the design of this ensemble configuration. There are also no perturbations designed for tropical cyclone structure or strength, or vortex relocation as is used in the NCEP ensemble.

3.2.3 NOGAPS ensemble with stochastic convection

One of the major issues with the NOGAPS ensemble discussed by McLay et al. (2008) is the ensemble variance is too small in the tropics and too large in the extratropics. This issue is partially due to the ensemble transform method not accounting for model error and uncertainties. Teixeira and Reynolds (2008) proposed a potential solution to this. They developed a stochastic convection scheme that accounts for uncertainties in the subgrid scale moist convective parameterization. In this scheme, probability density functions are used to constrain the random determination of future states based on the parameterization schemes themselves. Teixeira and Reynolds (2008) showed that the stochastic convection produces much greater ensemble variance in the tropics than the

ensemble with only the initial-condition perturbations, mainly because deep convection is common in the tropics. By blending with the ensemble transform perturbation method in a cycling scheme where the initial perturbations are based on previous short-term forecasts, the stochastic convection partially compensates the too-weak variance in the tropics in the stand-alone ensemble transform scheme (Reynolds et al. 2008).

Teixeira and Reynolds (2008) describe the stochastic convection scheme formulation, which we summarize here. Assuming that the standard deviation of a generic variable due to moist convection is proportional to its tendency, the following equation can be derived:

$$\frac{\phi_{conv}^{stoch} - \bar{\phi}}{\Delta t} = (1 + \eta\beta)\left(\frac{\Delta\bar{\phi}}{\Delta t}\right)_{conv} \tag{1}$$

where ϕ_{conv}^{stoch} is the stochastic value of the variable after convection, $\bar{\phi}$ is the mean value before the moist convection parameterization, $\bar{\phi}_{conv}$ is the mean value after convection, β is a constant of proportionality (which can be estimated from cloud resolving model studies), and η is a normally distributed stochastic variable. This stochastic component only affects moist convection with an assumed normal probability distribution function. Horizontal and temporal correlations are neglected, while the vertical correlation of the perturbations is determined by a single random number per column at each time step. The stochastic method is applied to wind and temperature variables only. Moisture is perturbed such that relative humidity remains constant. It should also be noted that the nonperturbed control member also has stochastic convection in this configuration.

3.3 Tracking methods

Similar to the NCEP cases, in this study we use only the forecasts out to 120 h from 0000 UTC. The ensemble forecasts are available at T239L30 resolution. Each investigation begins at least 60 h before tropical depression designation and continues at least 42 h after the system becomes a tropical depression.

Forecasts are compared between the ensemble forecasts with and without stochastic convection. The control ensemble (CTRL) is the NOGAPS ensemble configured with only ensemble transform initial perturbations. The stochastic method is included in the other set of NOGAPS ensemble forecasts (STO, hereafter) with $\beta = 1$. To prevent confusion, the control members (those without initial perturbations, both with and without stochastic convection) will be referred to simply as "member 0."

The same tracking techniques that were developed and used for the NCEP ensemble are used here: manual tracking of vorticity, geopotential height, and wind vectors at 850 hPa. This method continued to prove to be the most effective way to track the tropical systems. To maintain consistency, the same criteria for genesis (vorticity greater or equal to 7.5×10^{-5}, two closed height lines within 5 degrees at a 10 m interval, and closed circulation in the wind field) were also used for the NOGAPS ensemble. Vortex-like structure will be assigned to systems that meet only one or two of those criteria. A forecast will be labeled as "dissipation" if genesis criteria are met but were not maintained for more than 48 h. As shown later, using the same criteria verified as an appropriate metric to represent cyclone genesis.

3.4 Case studies

3.4.1 Nuri

Nuri developed from a convective complex with favorable upper level outflow which developed in the base of an upper trough well to the east of Guam on 11 Aug. It remained an open wave with some low level convergence and disorganized convection. The system continued to track westward with intermittent convection and limited development due to unfavorable upper level conditions. On 15 Aug, convection increased significantly under an upper-level high and low shear. A low-level cyclonic circulation (LLCC) developed on 16 Aug with good convective organization and low shear. The system had a distinct boundary with the upper trough to the north, limiting outflow. Another vortex to the south prevented a symmetric core from developing. There was low shear overhead but high values surrounded the system. High ocean heat content (OHC) existed in the system's path. The track was affected by a subtropical ridge, with which there was a lot of uncertainty. This disturbance rapidly developed from tropical depression (0000 UTC 17 Aug) to tropical storm (1200 UTC 17 Aug) to typhoon (1200 UTC 18 Aug). Some weakening occurred as it crossed the Philippine Islands, and it eventually made landfall near Hong Kong.

Five-day forecast tracking began with CTRL ensemble on 14 Aug 2008 (72 h before Nuri developed into tropical depression) when the vorticity maximum was located several hundred km to the east-northeast of Guam. While the forecasts from this day maintained the vorticity center, no members suggested cyclogenesis. Most forecast tracks followed a due west trajectory, although the forecast movement was slow when compared to the observed track (Fig. 16a). For forecasts begun on 15 Aug, a majority of

the forecast tracks still correctly showed westward movement, but storm motion was slower than observed. Ensemble spread was slightly larger compared to the previous forecast (Fig. 16b). While the number of members predicting genesis only increased by a few members, there were more forecasts for weak, broad circulations. The forecasts from 16 Aug showed significant improvement, with all 33 members predicting genesis. However, the forecast speed was still too slow, (Fig. 16c). In addition, some members began to predict northward curvature, and the ensemble spread was smaller than in previous forecasts. A similar trend continued in the forecasts from 17 Aug, with all members producing a slower speed than observed (Fig. 16d). More members also began to show northward curvature. In forecasts from 18 Aug, the members consistently maintained storm structure from the forecasts, but all members erroneously predicted northward curvature east of Taiwan around forecast hour 72 (Fig. 16e). Spread was small while the forecasts closely following the best track in the first stages of the forecast. A small increase in spread occurred after curvature was predicted. All of the forecasts from 19 Aug also predicted recurvature and eventual extratropical transition (ET; Fig. 16f). Initial strength was accurate, but all members predicted the cyclone to open into a trough as it transitioned.

For Nuri, stochastic convection only made minor improvements compared to the CTRL forecasts. Overall, there was slightly more ensemble spread in STO, however, the cyclone movement remained too slow. The ensemble mean and member 0 were very similar for CTRL and STO in forecasts from 14 Aug, with no significantly deviating members (Fig. 17a). For forecasts from 15 Aug, the spread was larger in STO than CTRL in the earlier forecast stages, but was similar in the later forecast stages. Some STO

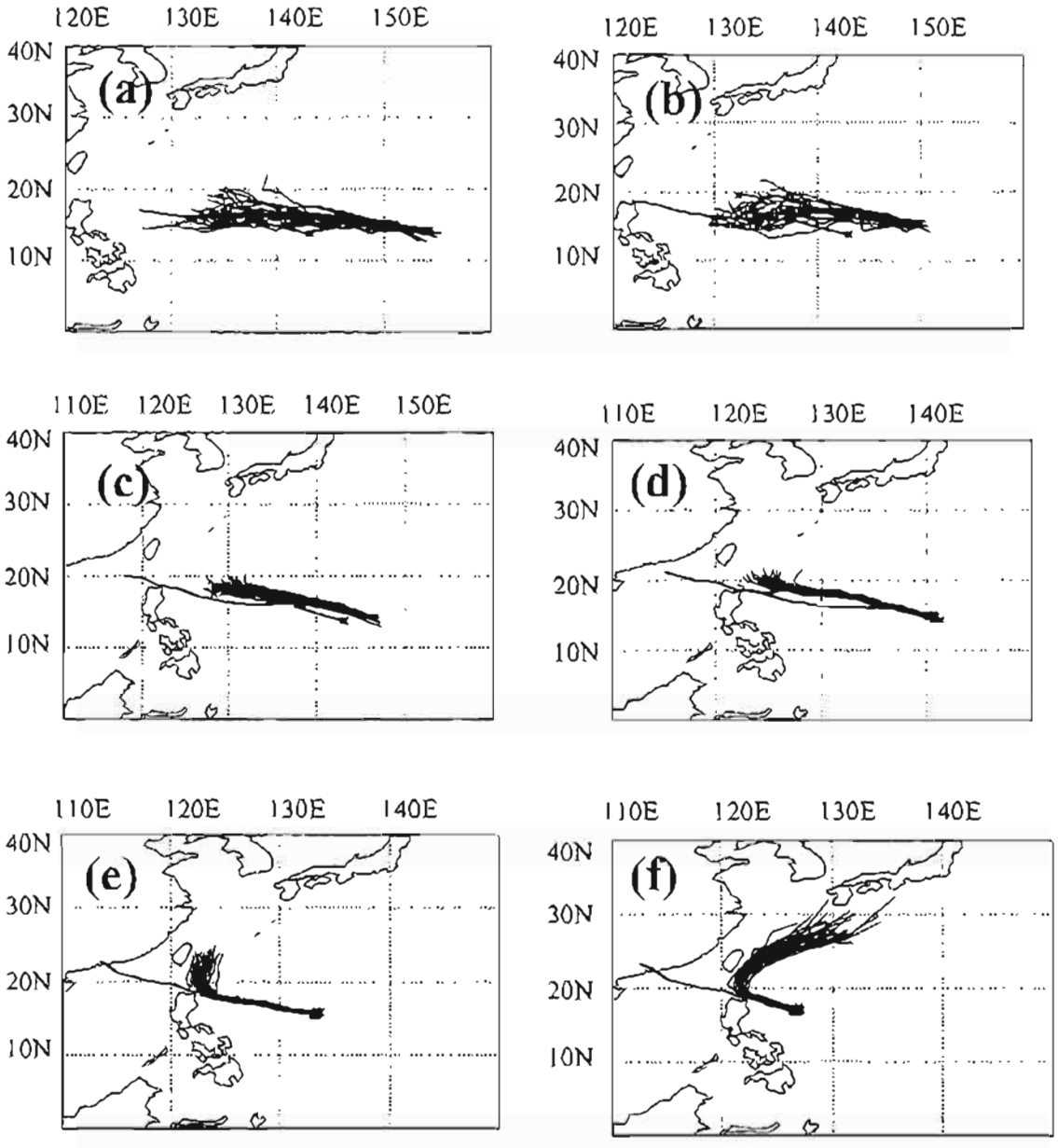


Figure 16: The tracks of 0-120 h CTRL ensemble forecast for Typhoon Nuri (purple thin lines) from 0000 UTC 14-19 Aug 2008 (corresponding to Fig. 16 a-f, respectively), compared with the corresponding "member 0" (blue line) and JTWC best track (thick black line). Red dashed lines denote the ensemble mean. An "X" designates the forecasted genesis.

members predicted faster movement (which pulled the ensemble mean farther west), although none captured the speed of the actual cyclone (Fig. 17b). While the spread was greater in STO on 16 Aug, all members remained tightly clustered with slow movement to the north of the actual track (Fig. 17c). In forecasts from 17 Aug, STO tracks demonstrated slightly more spread in early stages of the forecast, but were similar to CTRL in the later stages (Figs. 16d and 17d). All members remained north of the actual track. In the 18 Aug forecasts, there was little additional spread in STO until after the northward curvature. The STO tracks averaged a slightly more westward component to the curvature, but none followed the noncurving actual track (Fig. 17e). In forecasts from 19 Aug, there were more STO members that predicted rapid acceleration and ET after curvature. The spread was also greatest on this day, but all members continued to predict recurvature (Fig. 17f). In terms of intensity forecasts, genesis predictability improved for the 14 Aug forecast, while it worsened for the 16 and 17 Aug forecasts. All members showed reasonable storm structure in the postgenesis phase, although forecasts were affected by terrain interactions over Taiwan and the subsequent ET.

3.4.2 Sinlaku

On 1 Sep, intermittent convection northeast of Guam developed an interesting comma shape. The next few days, convection remained scattered and disorganized. Shear then eased and the system moved into an area of southwesterly flow. By 7 Sep, the system combined with another vorticity maximum, and enhanced convection and a strengthening low-level circulation were noted from satellite imagery (not shown). The system became a tropical depression at 1200 UTC 08 Sep and was immediately upgraded to Tropical Storm Sinlaku at 1800 UTC. Conditions became favorable with low shear and

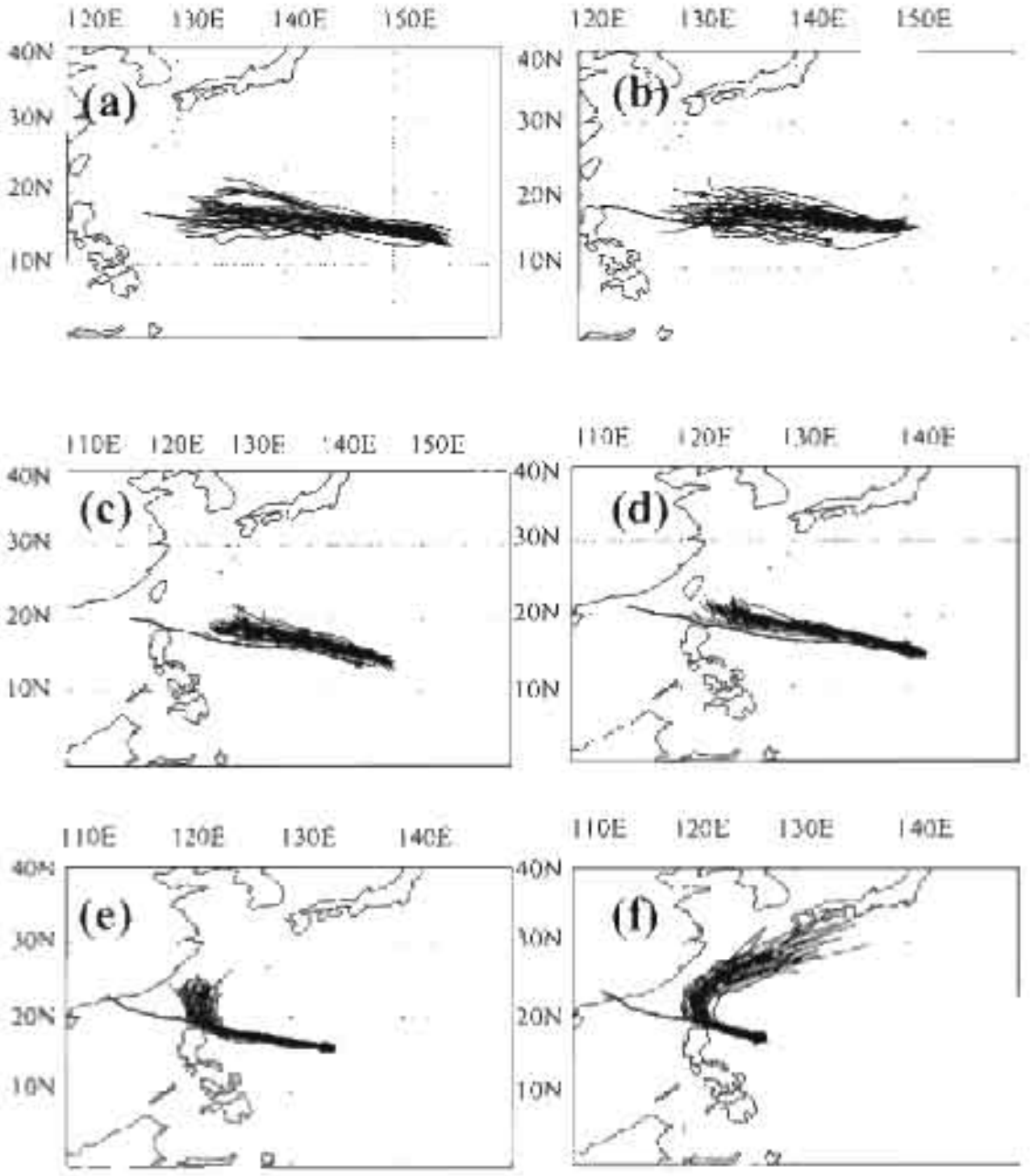


Figure 17: The tracks of 0-120 h STO ensemble forecast for Typhoon Nuri (purple thin lines) from 0000 UTC 14-19 Aug 2008 (corresponding to Fig. 17 a-f, respectively), compared with the corresponding "member 0" (blue line) and JTWC best track (thick black line). Red dashed lines denote the ensemble mean. An "X" designates the forecasted genesis.

high OHC. Sinlaku rapidly intensified to typhoon status by 1200 UTC 09 Sep. Eye wall replacement occurred on 11 Sep, although the cyclone was battling a weak ridge to the south and a strong ridge to the east that would determine its eventual track. The cyclone proceeded northwest toward Taiwan with increased shear on its northerly side due to an approaching upper trough. As the cyclone encountered Taiwan and lower OHC, it weakened significantly. When the cyclone curved northward, dry air entrained from the west, although radar images continued to show tight circulation and new convection on the southeast side (not shown). Weakening continued through 16 Sep due to westerly shear exposing the LLCC. The cyclone intensified 17-19 Sep as the subtropical ridge broke and allowed the cyclone to move northeastward over the warm Kuroshio Current, but movement occurred under moderate shear, which tilted the vortex. A warm core structure was also maintained during these days. By 20 and 21 Sep, ET fully occurred as the cyclone was picked up in the mid-latitude westerlies and moved rapidly to the east.

Five-day forecast tracking started with CTRL forecasts from 6 Sep 2008, 60 h before the tropical system became a tropical depression. While all members tracked the cyclone westward, there was some disagreement as the cyclone stalled near 20°N 155°E. While the actual cyclone moved northward through this region, all the members depicted a track north of the initial genesis location, and few showed the observed northward turn (Fig. 18a). Some members predicted broad circulations, but none met the criteria for genesis. From the forecasts started on 7 Sep, most members were predicting broad and weak circulations or complete dissipation. The forecasts were rather diverse and complex, with many members showing westward movement initially, followed by backtracking to the east (Fig. 18b). Few members tracked the disturbance over the eventual genesis

region or showed the observed northward turn. The trend changed for the forecasts from 8 Sep. Fig. 18c shows that the ensemble spread was reduced, and most tracks followed a north then northeast trajectory, but much faster than the actual track. Probability of cyclone genesis remained low. Most members in the forecast from 9 Sep predicted rapid north-northeast movement and ensuing ET (Fig. 18d). However, one member did predict a track that was similar to the actual track, which was further west. All members depicted accurate initial strength, but were split between deepening the system or forecasting it to remain weak and open into a trough. The forecasts from 10 Sep significantly increased the ensemble diversity. While some members predicted rapid northeast movement and ET, other predicted slower westward movement (Fig. 18e). Genesis criteria were met initially by all members, but most predicted weakening as storm-structure evolved. A reasonable cone of uncertainty was produced from the forecasts begun on 11 Sep (Fig. 18f). Some members predicted rapid ET, while others took a slower course that was closer to the actual track. The storm intensity forecasts were also dependent on these outcomes, with interactions due to both terrain in Taiwan and baroclinicity in the midlatitudes.

STO primarily added more spread to the tracks of pregenesis forecasts (Fig. 19). Overall the westward stall and retreat track remained similar. STO tracks were especially erratic in the forecasts started on 6 Sep, although the mean was still similar to the CTRL ensemble mean (Figs. 18a and 19a). The forecasted STO tracks from 7 Sep were much more diverse than CTRL, although a lack of consensus provided little added value in the ensemble mean from STO ensemble (Figs. 18b and 19b). For forecasts from 8 Sep, the STO mean showed northeast movement much like the CTRL, although a westward shift

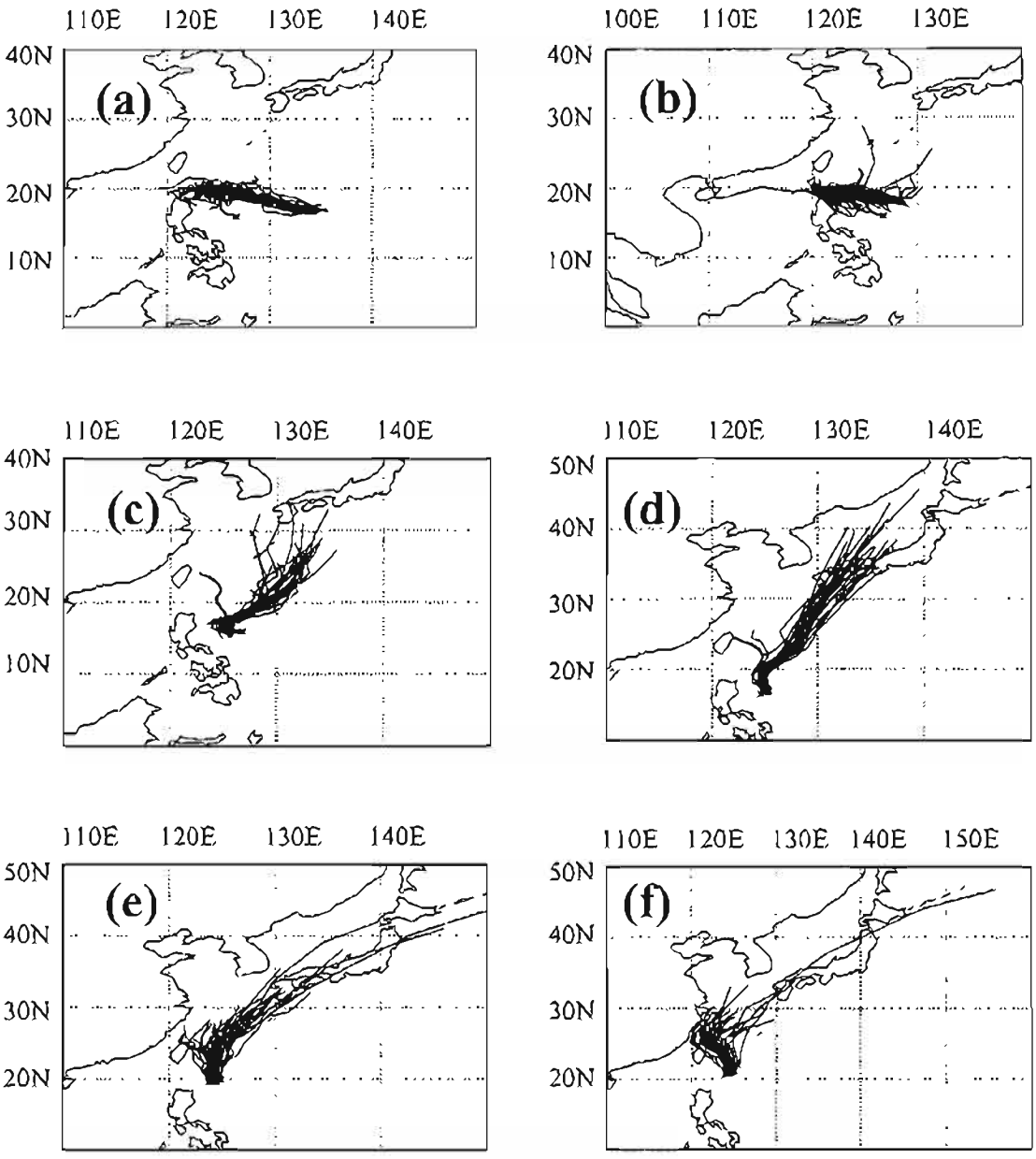


Figure 18: The tracks of 0-120 h CTRL ensemble forecast for Typhoon Sinlaku from 0000 UTC 6-11 Sep 2008 (corresponding to Fig. 18 a-f, respectively), compared with the JTWC best track.

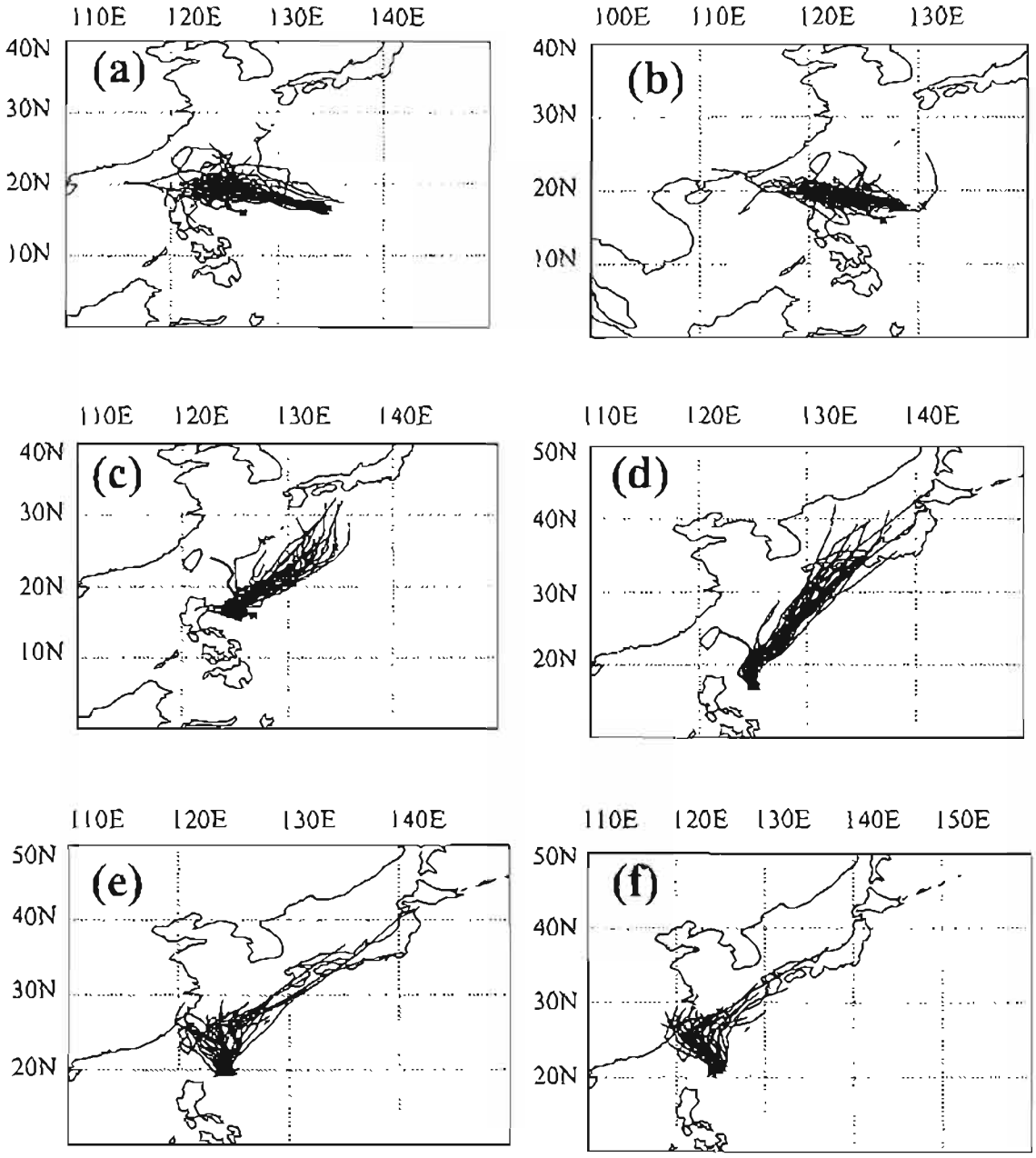


Figure 19: The tracks of 0-120 h STO ensemble forecast for Typhoon Sinlaku from 0000 UTC 6-11 Sep 2008 (corresponding to Fig. 19 a-f, respectively), compared with the JTWC best track.

was noted in the ensemble envelope. The STO mean was also somewhat slower than the CTRL mean (Figs. 18c and 19c). On 9 Sep, the general trend in the forecasts was for rapid northeast movement like CTRL, although the STO mean track was not as far to the northeast (Figs. 18d and 19d). From the forecasts on 10 Sep, there were fewer cases of rapid ET in STO than CTRL. As seen in Fig. 19e, members were evenly spread amongst the observed western track, the central ensemble mean, and the erroneous northeast track (which was not the case in CTRL). A similar trend continued in forecasts from 11 Sep with many STO tracks farther west than observed (Fig. 19f).

Regarding genesis predictability, there were fewer predictions of vortex development on 6 Sep, but there was a higher genesis rate in the forecasts from 7 Sep. The STO forecast from 8 Sep showed a significant increase in genesis predictability, although the total number of members predicting genesis (11) was low considering the forecast was made only 12 h prior to genesis. Genesis criteria were met for all postgenesis forecasts. However, intensity forecasts depended on whether the cyclone was forecast to move toward Japan and strengthen (mainly forecasts from 9 Sep) or interact with Taiwan's terrain and weaken (forecasts from 10 and 11 Sep).

3.4.3 Hagupit

Disorganized convection formed on the southeast side of a tropical upper-tropospheric trough (TUTT) east of Guam on 12 Sep, under upper-level divergence but strong shear. The system continued to battle an unfavorable upper level environment. On 16 Sep, convection intensified with a broad surface circulation center forming, although the mid-level circulation center was located further south. The system encountered shear the next day and subsequently weakened. Convection increased in aerial extent on 18 Sep

and the system became a tropical depression at 1800 UTC, but the circulation remained under strong shear. Tropical Storm Hagupit was designated on 19 Sep at 0600 UTC under weakened shear, with the deepest convection to the southeast (convergence region). A subtropical ridge inhibited poleward outflow. However, a trough digging over Japan helped enhance the outflow. Convection continued to be concentrated on the southern section as the system slowly intensified and moved into a low shear/high OHC environment. Considerable intensification occurred by 21 Sep (category 3 typhoon), with further intensification afterward. Some weakening occurred over lower OHC before the cyclone made landfall on 24 Sep in China.

Model evaluation of the 120 h forecasts began on 16 Sep, 66 h before genesis. The genesis rate was high for these forecasts, with 23 members predicting cyclone development. However, track forecasts were poor, with all members turning the system northwestward well east of the actual genesis location (Fig. 20a). A small cone of uncertainty was depicted by the track forecasts. From the forecasts on 17 Sep, the ensemble track trend was for a northward turn, which occurred east of the actual genesis location (Fig. 20b). There was also a decrease in the number of members predicting genesis, with some predicting a transition into a mid-latitude trough. Spread amongst the members was small and uniform over the forecast period. The forecasts from 18 Sep continued to show an immediate turn to the north (Fig. 20c). Genesis predictability reduced further, with only four members showing development. Members predicting nongenesis were split between a broad, weak system, and one that merged into the mid-latitudes. The track forecasts remained tightly clustered. While forecasts initialized 19 Aug forecasts forecasted more northwesterly tracks, all members remained north of the

actual track (Fig. 20d). The cone of uncertainty was more realistic, but at no point encompassed the actual track. Comparatively, the ensemble forecasts were also much slower than the actual track. With initialization 6 h before tropical storm designation (and 6 h after tropical depression designation), all 33 members met genesis criteria and sustained development. Again, forecasts initialized on 20 Sep remained on a northwest track that was further north than observed (Fig. 20e). Spread increased toward the end of the forecast period, with some members showing recurvature, and others depicting a westerly turn. In forecasts from 21 Sep, the ensemble members became more aligned with the actual track, with speed and direction nearly identical (Fig. 20f). Spread increased with forecast time. Intensity forecasts initialized on both 20 and 21 Sep were initially strong but weakened with time, mainly due to interactions with the high terrain located in Taiwan.

For both forecasts from 16 and 17 Sep, track spread increased in STO, but the envelope of ensemble tracks still did not encompass the best track (Figs. 21 a-b). On the other hand, more STO members followed a track with a more westward component than CRTL members did. However, several STO members forecasted more northerly progress than any of the CTRL. Additionally, the probability of genesis increased both days over that of CTRL. Spread increased dramatically in the forecasts from 18 Sep, with a few members showing westward movement, although tracks were much slower than the actual cyclone (Fig. 21c). The STO mean was only slightly west of the CTRL mean, and STO members that predicted recurvature were somewhat faster than the CTRL members. The genesis predictability was higher with STO than the CTRL, but still decreased dramatically from the rate on 17 Sep. The forecasts from 19 Sep also showed a large

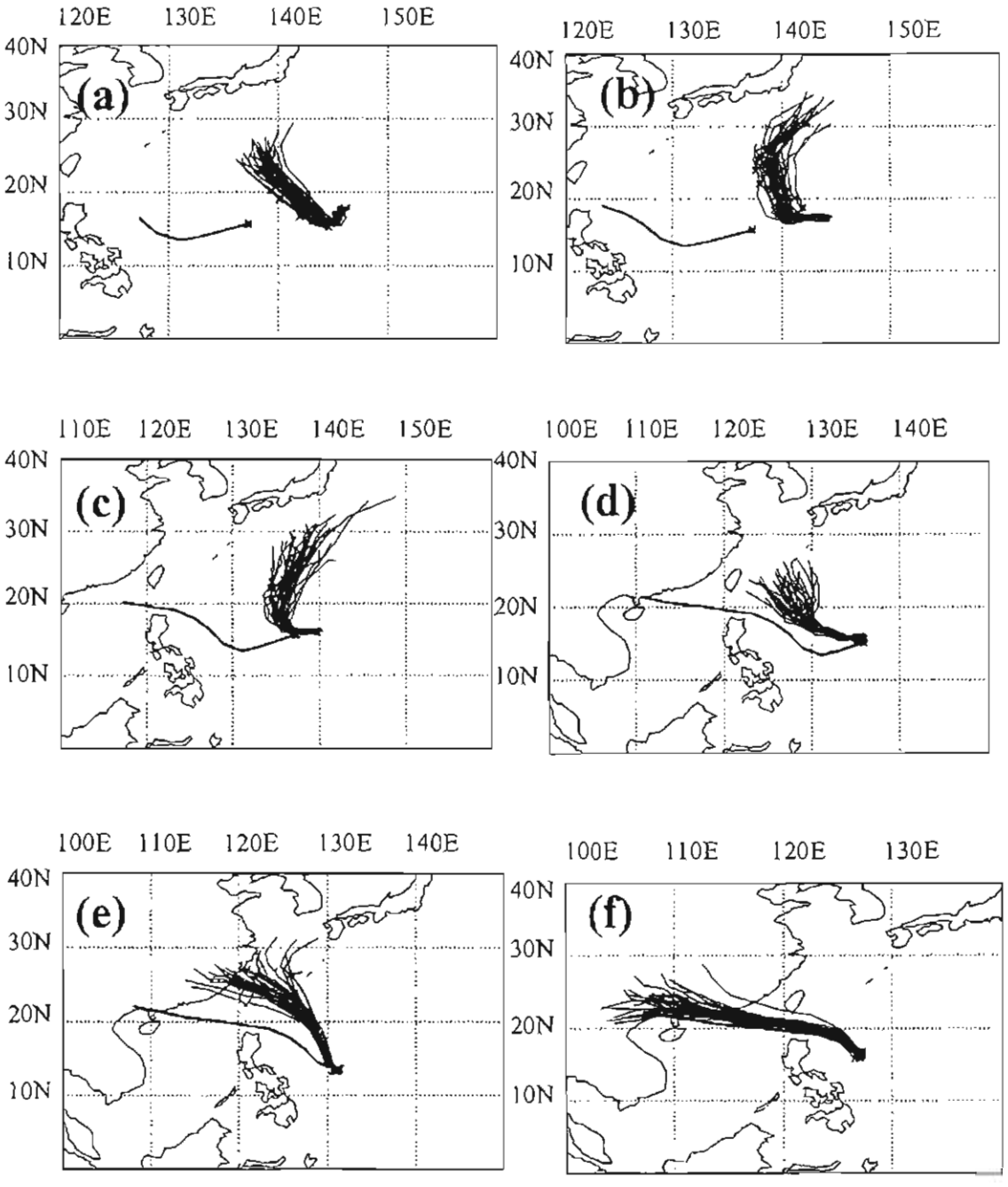


Figure 20: The tracks of 0-120 h CTRL ensemble forecast for Typhoon Hagupit from 0000 UTC 16-21 Sep 2008 (corresponding to Fig. 20 a-f, respectively), compared with the JTWC best track.

amount of spread, with some members recurving the cyclone and others moving it to the west (Fig. 21d). Overall, most members were north of the actual track and all were too slow, although the STO forecasts were on average slightly faster than CTRL. The number of STO members meeting genesis criteria was slightly lower than CTRL due to a few cases of premature dissipation. All members were too far north in the forecast tracks from 20 Sep, although spread was increased over CTRL (Figs. 20e and 21e). In forecasts from 21 Sep, spread was increased during the first half of the forecast period, but was similar to CTRL for the latter half (Figs. 20f and 21f). Much like CTRL, STO forecasts from both 20 and 21 Sep initialized intensity correctly, but forecast weakening as the cyclone crossed Taiwan and interacted with the China mainland.

3.4.4 Jangmi

Jangmi originated as an area of intense convection developed on the south side of a TUTT cell east of Guam on 16 Sep. Intermittent and scattered convection continued through 22 Sep with high shear and an upper-level trough to the north. Convection became more organized on 23 Sep with an LLCC evident, low shear, and good upper-level outflow. The JTWC designated the system a tropical depression at 1800 UTC, followed by tropical storm status at 0000 UTC 24 Sep. The upper level pattern, OHC, and surface flow were all favorable for development. Deep convective bands were evident near the center with excellent flow to the southeast (TUTT inhibited flow to the north). West-northwest motion continued as the cyclone punched into a mid-level ridge. At 0600 UTC 25 Sep, the cyclone reached typhoon status with a well-developed eye and continued intensification. Under similar conditions and gradual strengthening on 26 Sep, forward motion slowed. Rapid intensification into super typhoon status occurred on 27

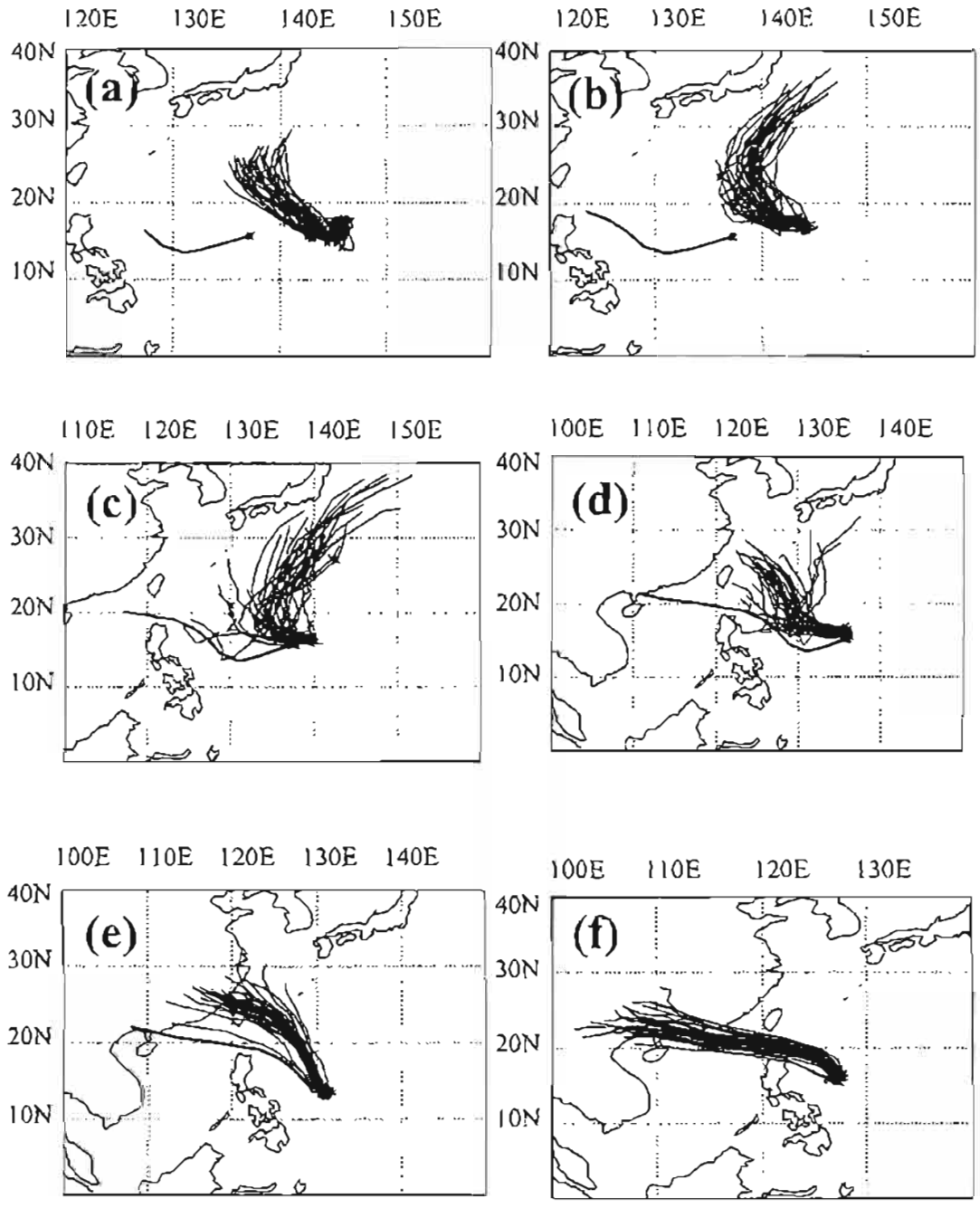


Figure 21: The tracks of 0-120 h STO ensemble forecast for Typhoon Hagupit from 0000 UTC 16-21 Sep 2008 (corresponding to Fig. 21 a-f, respectively), compared with the JTWC best track.

Sep. However, the cyclone weakened prior to crossing Taiwan on 28 Sep. Over 500 mm of rain fell over parts of Taiwan. The cyclone lost its convective structure on 29 Sep as it encountered high shear. On 30 Sep, the cyclone fully underwent ET with strong upper-level southwest flow shearing convection away. The cyclone was ultimately swept away into a frontal zone with minimal downstream development.

For the first day of tracking, 21 Sep (66 h before genesis), the 120 h CTRL forecast tracks were consistent with westward movement, but a majority of the ensemble members produced tracks that were farther south and slower than observed (Fig. 22a). No members suggested that this wave would develop into a cyclone. Spread increased little with time. While the tracks from 22 Sep mirrored the northwesterly turn of the actual cyclone, most member tracks were too far south and too slow (Fig. 22b). Ensemble spread was similar to the previous day. Development into a tropical cyclone was not shown by any member, with most cases showing a large, weak circulation. In the forecasts from 23 Sep, the tracks remained slow and slightly south of observed, with very little spread (Fig. 22c). Members were nearly evenly split between forecasts of cyclone development or continued weak circulation. The accuracy of the track forecasts beginning 24 Sep deteriorated in the latter stages of the forecast period, showing a westerly turn, whereas the actual cyclone continued northwest (Fig. 22d). Spread amongst members was small. All members initialized intensity accurately and predicted moderate strengthening. A northwesterly trajectory closer to the actual track was predicted by the forecast from 25 Sep. However, the tracks from most members slowed and drifted westward by the end of the period (Fig. 22e). Only slight to moderate strengthening was forecast, with minimal ensemble spread. The westward drift of the

previous day continued in the forecasts from 26 Sep. A majority of the members predicted landfall in China (Fig. 22f). The members that predicted a northward curvature were much slower than observed and did not reach the ET stage. Outside of the few northward turns, spread was small, with reasonable intensity forecasts.

There was a large amount of spread in the STO forecast from 21 Sep. Even though several members encompassed the actual track, all forecasts were too slow (Fig. 23a). The STO ensemble mean was farther north than the CTRL mean. These trends continued for forecasts from 22 Sep, with a few STO members along the correct track and speed, along with a slightly more northward ensemble mean (Fig. 23b). The slowing trend reappeared in the forecasts from 23 Sep, but spread was still greater than CTRL (Figs. 22c and 23c). Some STO members forecast the system to be absorbed into a large gyre, with subsequent dissipation. Once again, the STO mean was slightly more northward and closer to observed. Although more spread and a northerly trend existed in the STO forecasts from 24 Sep, the ensemble forecasts generally agreed on a westerly track forecast (Fig. 23d). An interesting trend appeared in the forecasts from both 23 and 24 Sep, where some members predicted an interaction between the original vorticity maximum and one farther west. Sometimes this forecast consisted of a simple merger, while other times a Fujiwhara effect took place (Fujiwhara 1923). In some cases, the original vorticity maximum would completely dissipate, while the western one dominated. Even though the physicality of these scenarios may be debated, none of them verified. On 25 Sep, a few STO members predicted recurvature, but it occurred too early in the track. The other members continued the westerly track (Fig. 23e). This day was also the first time a majority of the members showed rapid deepening. A more split solution evolved in

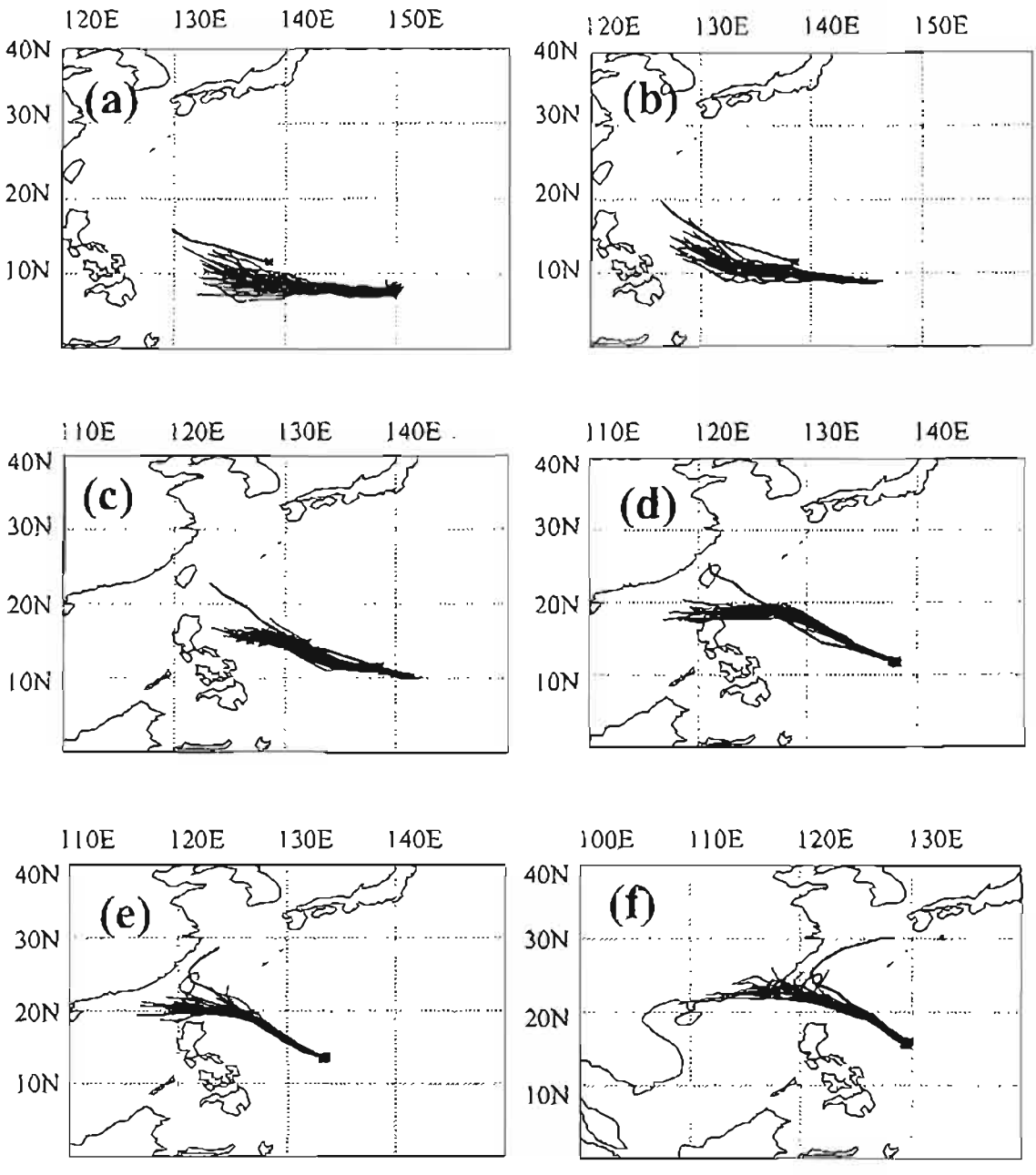


Figure 22: The tracks of 0-120 h CTRL ensemble forecast for Typhoon Jangmi from 0000 UTC 21-26 Sep 2008 (corresponding to Fig. 22 a-f, respectively), compared with the JTWC best track.

the STO forecasts from 26 Sep, with some members predicting westward movement and others predicting recurvature (Fig. 23f). However, no members exactly matched the observed track. The number of members predicting genesis was consistently six to eight members more for STO than CTRL. Even then, only 18 members predicted genesis (with two vortex-like cases) for this forecast, which initialized 18 h prior to the JTWC tropical depression designation. The STO forecasts did prove to be inferior beginning on 24 Sep, with a few intensity initializations only meeting "vortex-like" criteria. For 25 and 26 Sep, STO forecasts initialized with appropriate intensity and predicted the cyclone to deepen until it made landfall.

3.4.5 Higos

A large region of convection developed on 22 Sep both north and south of the equator due to a westerly wind burst. Convection remained minimal and dry air was wrapped into the system, but a corridor of low shear stretched across the tropics. Organization improved on 25 Sep, especially at the low levels. Some consolidation occurred without aid from the monsoon trough, but Jangmi's outflow was nearby to inhibit development. Winds increased on 29 Sep with convective banding features evident. The northeasterly flow weakened over the South China Sea on 30 Sep, and Higos reached tropical storm status, but weakened again as it encountered the terrain of the Philippines. Even though shear weakened and the cyclone took a westerly turn, only marginal redevelopment occurred over the next few days. The cyclone took a northward hook over Hainan and made landfall in China on 4 Oct.

Model evaluation began 78 h before tropical depression designation at 0000 UTC 26 Sep. Five-day CTRL forecast tracks were fairly uniform with minimal spread initially.

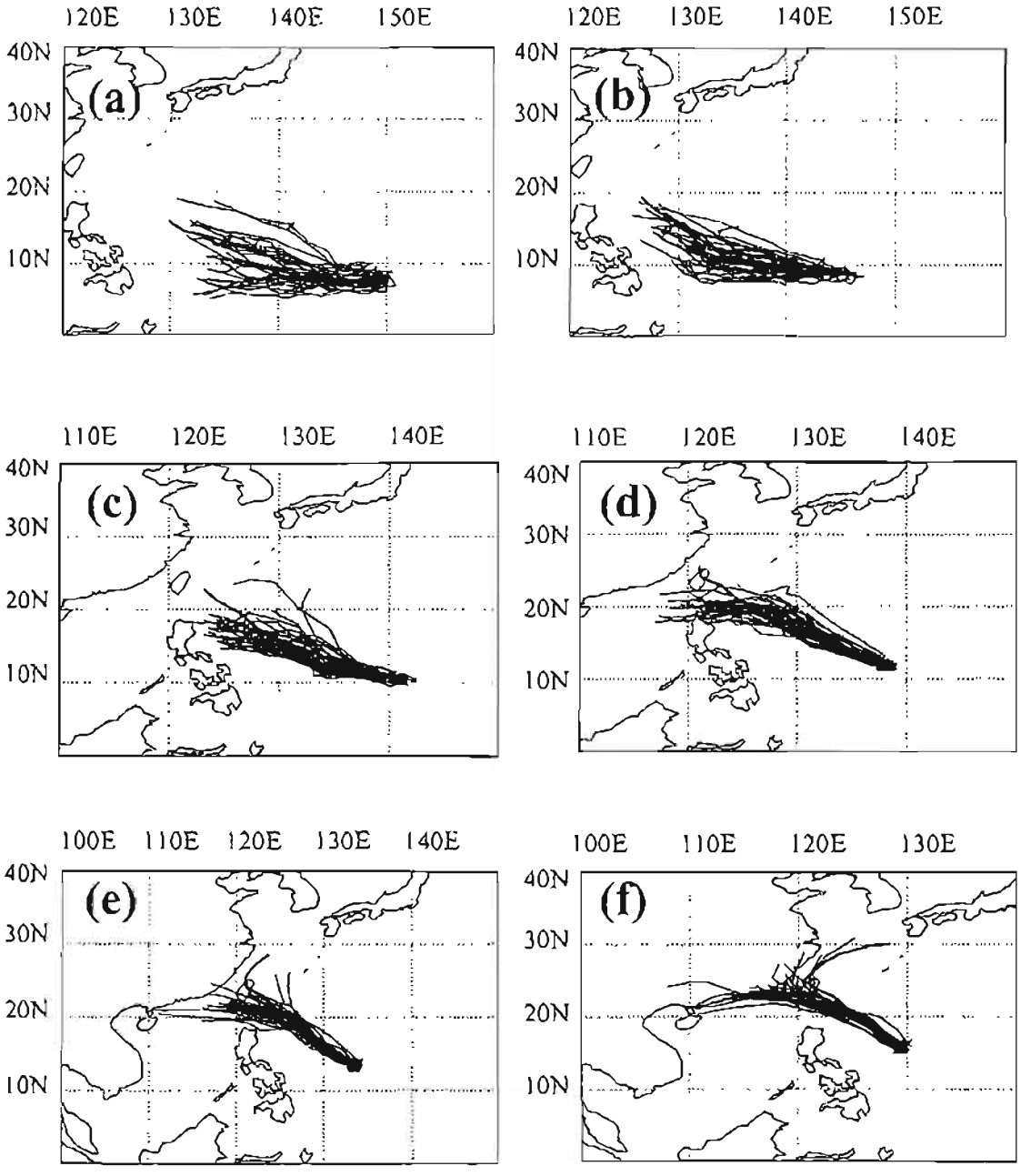


Figure 23: The tracks of 0-120 h STO ensemble forecast for Typhoon Jangmi from 0000 UTC 21-26 Sep 2008 (corresponding to Fig. 23 a-f, respectively), compared with the JTWC best track.

Ensemble spread increased late in the forecast period with some members predicting northward curvature and others maintaining a westerly course (Fig. 24a). However, almost all of the members predicted a track that was north of observed. Cyclone genesis was not predicted by any of the members, with many members forecasting elongation of the vorticity maximum and eventual dissipation. The forecasts from 27 Sep were much more ambitious for the storm's intensity. All members predicted at least vortex-like strength. However, all the members also predicted a northerly turn (staying north of the actual track for the entire forecast), as seen in Fig. 24b. Therefore, later in the forecast period, some members predicted that the circulation would open into a wave as it interacted with mid-latitude features. Ensemble spread was minimal until curvature was forecast. On 28 Sep, the forecast tracks were closer to observed initially, but most members continued to predict a northward turn (Fig. 24c). In addition, all of the ensemble members forecast systems that moved much slower than the observed. Intensity forecasts remained on the moderate side, with a majority of the members predicting vortex-like systems, most of which were forecast to later weaken substantially. Initially, tracks in the forecasts from 29 Sep closely followed the actual track, albeit with little spread (Fig. 24d). All members predicted a northward turn between 120° and 125° W, with slightly more ensemble spread. Intensity forecasts showed moderate development, although a majority of the members reached genesis criteria (with the remainder being vortex-like). All of the forecasts from 30 Sep initialized meeting genesis criteria, and while most maintained this intensity, some members forecast further strengthening. Forecast tracks were much closer to the actual track, but did not follow the exact cyclone path (Fig. 24e). Spread increased with time but was small overall. Ensemble tracks from 1 Oct predicted early recurvature,

but the general storm motion was close to that of observed (Fig. 24f). Spread increased with time, and the ensemble envelope included the actual track throughout the forecast period. Forecasts predicted similar intensity (some members showed slight weakening with subsequent redevelopment) throughout the forecast period. A majority of the members predicted landfall near the end of the period, closely coinciding with observed.

The STO forecast from 26 Sep had increased spread over the CTRL in the early stages of the forecast, but was similar in later stages (Figs. 24a and 25a). The STO mean showed a similar northward turn. For each respective member 0, STO took a southerly turn toward the actual track, opposite of the northerly turn of CTRL. While forecast intensity was generally weak, there were several members that predicted strengthening. For the forecasts from 27 Sep, the STO tracks had more spread and a more northwesterly course overall than CTRL (Figs. 24b and 25b). Speed was also slightly faster. In general, STO forecasts predicted a much stronger system, with double the number of genesis forecasts than CTRL. A similar trend continued in forecasts from 28 Sep, with the STO ensemble mean and envelope being slightly faster and to the northwest compared to CTRL (Figs. 24c and 25c). Intensity forecasts dramatically increased, with a majority of STO members reaching genesis criteria. In some cases, these forecasts for strengthening continued, with a major cyclone being prognosticated. The forecasts from 29 Sep also showed increased spread over CTRL (Figs. 24d and 25d). The northwest shift of the STO mean and ensemble envelope was also much more pronounced. However, all members still predicted northward curvature prematurely. While STO had only a few more members predicting genesis intensity, most of the members continued to predict a strengthening cyclone that would become large with time. There was more track spread

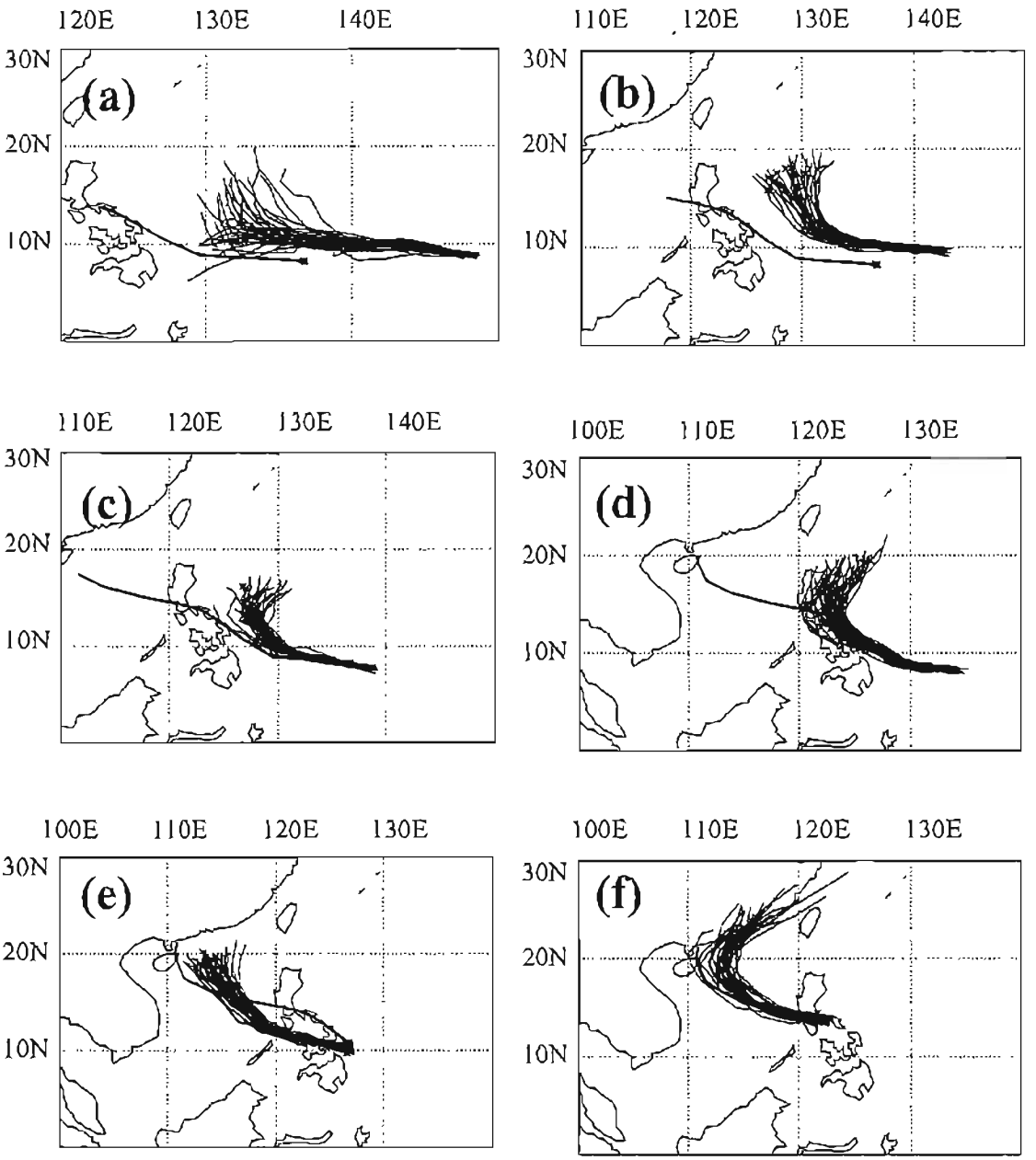


Figure 24: The tracks of 0-120 h CTRL ensemble forecast for Tropical Storm Higos from 0000 UTC 26 Sep-1 October 2008 (corresponding to Fig. 24 a-f, respectively), compared with the JTWC best track.

in forecasts from 30 Sep, but the STO mean remained similar to the CTRL mean (Figs. 24e and 25e). Intensity forecasts continued to be strong compared to both CTRL and observed, with a majority of the members predicting an intense cyclone. In addition to more spread, the STO track forecasts from 1 Oct were also slightly slower and farther to the west, with fewer members predicting rapid northeast movement (Fig. 25f). While not as ambitious as previous forecasts, many STO members continued to predict a strengthening cyclone.

3.4.6 TCS017

An area of convection that could be traced back to Tropical Storm Kika, which originated in the eastern North Pacific, drifted westward into the T-PARC/TCS-08 domain. Although the system weakened with time, shear in the region was weak. A weak LLCC was noted on 14 Aug, but convection remained poorly organized. Sea surface temperatures and upper-level divergence were favorable for tropical cyclone development, but complex interactions with the mid-to-upper levels hindered development. Circulation and convection remained weak for several days. Shear remained strong over the system. Convection was intermittently strong due to southwesterly winds feeding into the disturbance, but no LLCC developed. The system was eventually absorbed by the upper-level low and advected quickly north, dissipating by 25 Aug.

Tracking was chosen to begin on 19 Aug, when the system started to show signs of organization. In the forecasts from this day, the 120-h ensemble tracks had a medium amount of spread, with erratic cyclone tracks in the later stages of the forecasts (Fig. 26a). There were no forecasts that indicated cyclone development, and many forecasts

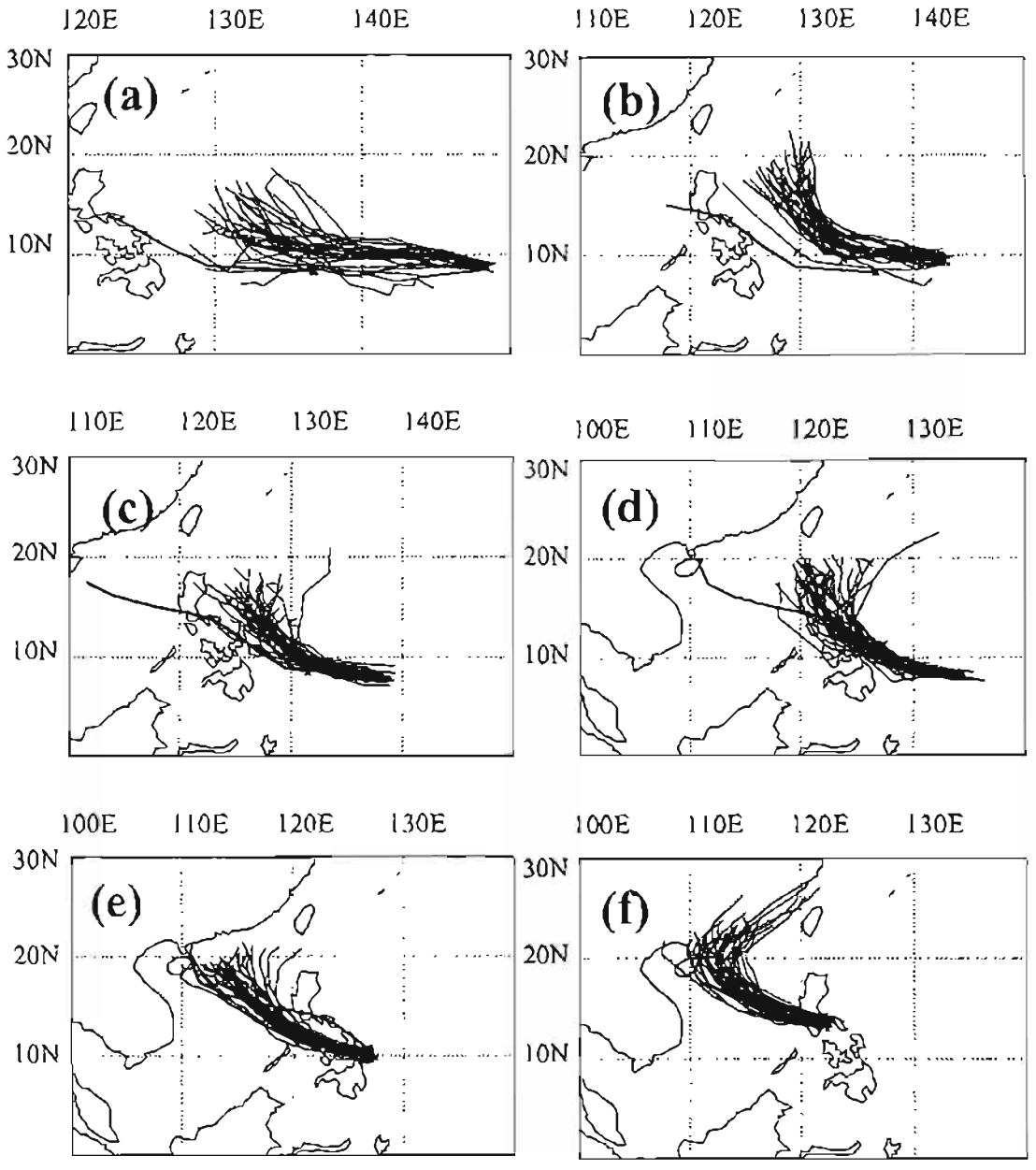


Figure 25: The tracks of 0-120 h STO ensemble forecast for Tropical Storm Higos from 0000 UTC 26 Sep-1 October 2008 (corresponding to Fig. 25 a-f, respectively), compared with the JTWC best track.

predicted dissipation or merger into the mid-latitudes. The ensemble mean for the CTRL forecasts from 20 Aug showed curvature to the northeast, but spread was large, with some members depicting northwesterly movement (Fig. 26b). Several members met vortex-like and genesis criteria, which were the only forecasts which suggested development. However, the system was still forecast to be absorbed into the mid-latitudes in the later stages. Forecasts consolidated slightly on 21 Aug, with most members forecasting a north to northeast track (Fig. 26c). Spread was still larger than in most of the well-defined cases. There were a few cases of vortex-like development, but most members forecast dissipation or merger into the midlatitudes. In the forecasts from 22 Aug, spread continued to be large as the model struggled with how the system would rotate around the subtropical gyre. The ensemble mean represented this consensus well, with a counter-clockwise track (Fig. 26d). There were more forecasts for the vorticity maximum to maintain its structure, but there were still only a few forecasts of vortex-like development. Additional forecasts were not studied due to the fact that the system was nearing the end of its life cycle.

With the high uncertainty and weak strength of TCS017, the STO ensemble forecasts provided less value over CTRL than for the named cyclones. For the forecasts from 19 Aug, there was slightly more spread in STO, and members generally forecasted slightly higher speed (Figs. 26a and 27a). A few more members predicted further development than CTRL. In the forecasts from 20 Aug, spread noticeably increased in the early stages of the forecast, but was more similar to CTRL in later stages (Figs. 26b and 27b). This forecast provided one of the stronger examples of the ensemble mean shifting between CTRL and STO. For the STO mean, the curvature associated with the

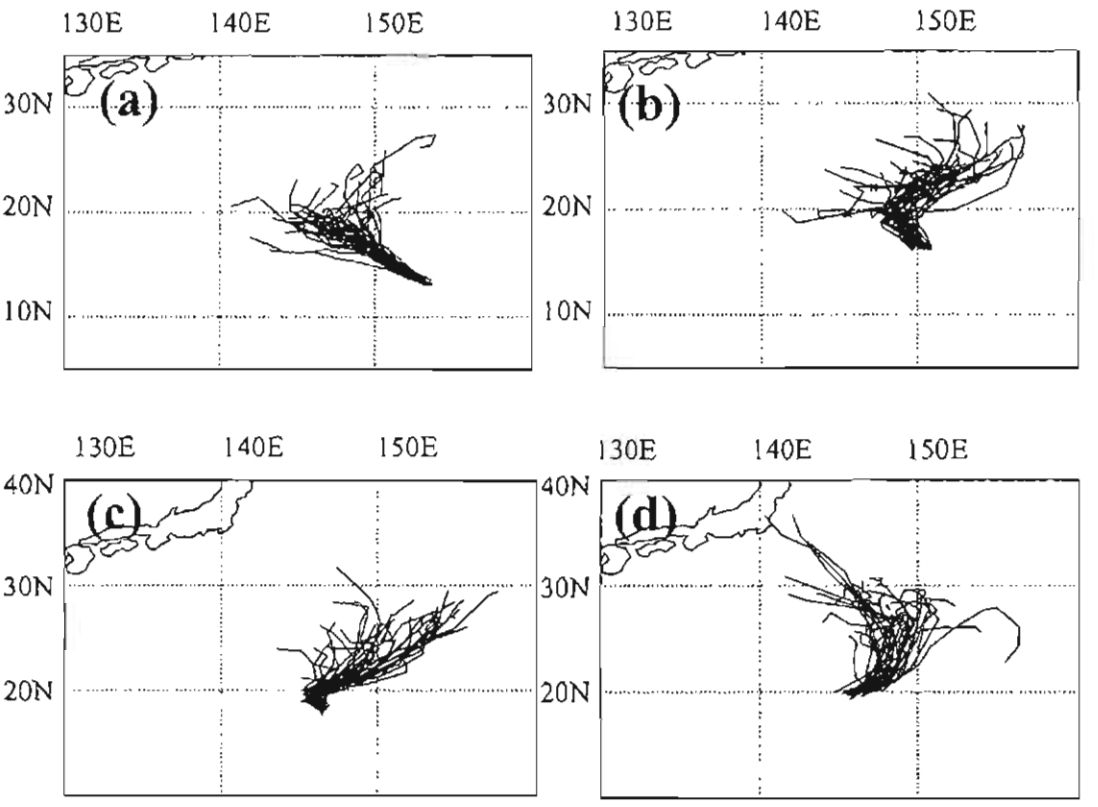


Figure 26: The tracks of 0-120 h CTRL ensemble forecast for TCS017 from 0000 UTC 19-22 Aug 2008 (corresponding to Fig. 26 a-d, respectively), compared with the JTWC best track.

gyre occurred earlier (farther west) in the track and was less pronounced than in CTRL. This STO forecast also contained a higher number of members forecasting genesis criteria to be reached. The STO forecasts from 21 Aug demonstrated a large amount of spread. Ensemble member tracks were divergent from the start of the forecasts (Fig. 27c). While the STO ensemble mean showed curvature around the gyre like CTRL, it was located farther west and had a slightly higher speed. On average, STO intensity forecasts were similar to CTRL. Similarly, STO forecasts from 22 Aug also produced divergent tracks, although most of the spread was concentrated later in the forecast period (Fig. 27d). The STO mean demonstrated similar curvature to the CTRL mean did, but was slightly more progressive with its speed. Like the CTRL, there were many forecasts for the vorticity maximum to maintain its strength, but only a few forecasts for vortex-like development.

3.4.7 TCS030

An area of organized convection developed south-southeast of Guam on 28 Aug. Shear was moderate, with lower values to the west. The key to development was related to the amplitude of the upper-level trough digging to the south (almost to 10°N). Convection became weaker with time as shear increased, although conditions aloft were favorable for development. The precise location of the upper-level trough led to a weaker system (divergence not co-located with convection), and conditions for development lessened as the system lost the upper-level support. Easterly low-level winds remained persistent, supporting the possibility for weak development. The system continued to track westward as a weak wave with minimal convection and weak low-level features. By 6 Sep, the system became absorbed in southwesterly flow in the South China Sea.

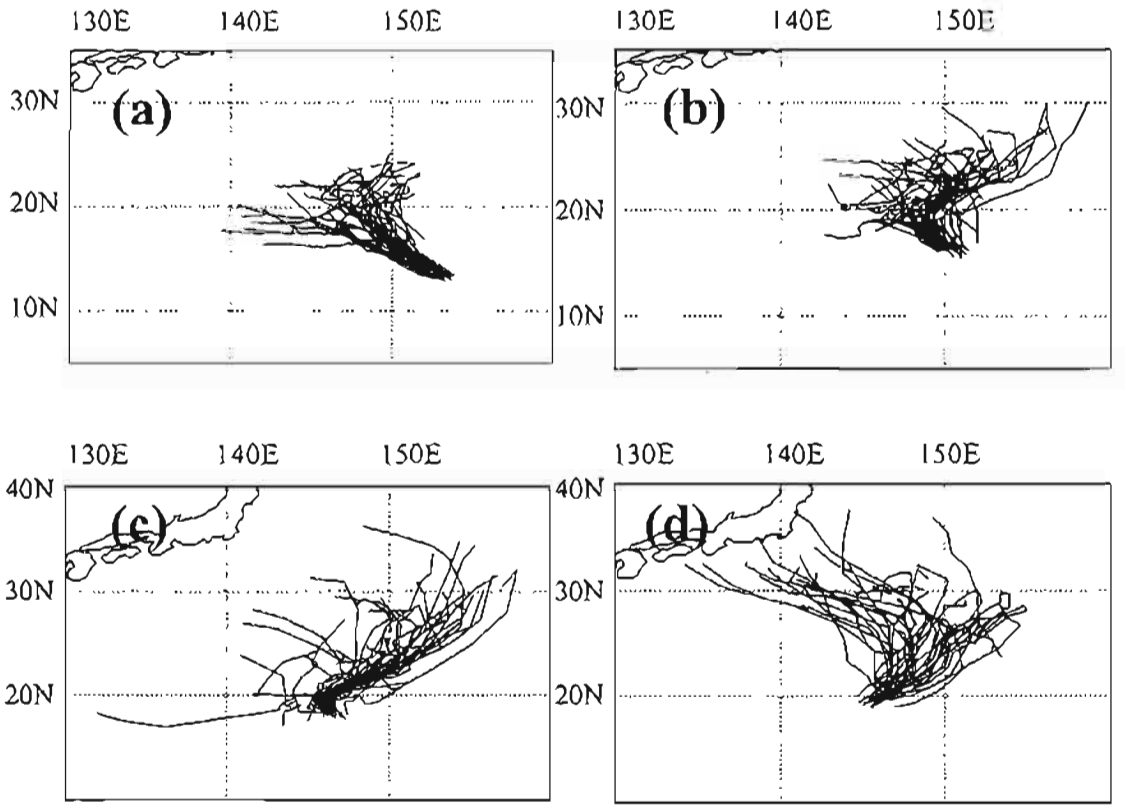


Figure 27: The tracks of 0-120 h STO ensemble forecast for TCS017 from 0000 UTC 19-22 Aug 2008 (corresponding to Fig. 27 a-d, respectively), compared with the JTWC best track.

Five-day forecast tracks began at 0000 UTC 31 Aug, when the NOGAPS began to prefer TCS030 for development over a nearby vorticity maximum. Most of the CTRL ensemble forecasts on this day predicted that the system would remain broad and weak well into the forecast period (Fig. 28a). However, gradual strengthening was forecast to occur, when the majority of the members indicated vortex-like development, but none met genesis criteria. The forecasts from 1 Sep also developed a broad vorticity maximum along a nearly-westerly track (Fig. 28b). Spread increased over the previous day, with some members suggesting a more northerly component to the track. The number of vortex-like cases decreased, and forecasts for cyclone genesis remained nonexistent. A westerly track forecast continued on 2 Sep, with minimal spread at the beginning of the forecast and increasing with time (Fig. 28c). Intensity forecasts continued a weakening trend, with most members developing a very broad and weak circulation, but not meeting vortex or genesis criteria. Starting from 3 Sep, track forecasts showed a northwesterly turn in the later part of the period (Fig. 28d). Spread amongst the tracks increased when this turn occurred, but was small before then. Most members kept the system's intensity weak, with only a few cases of vortex-like development. Prospects for development would continue to decline after this point, so no further forecasts were tracked.

The STO forecasts from 31 Aug were different than CTRL, both in track and intensity. While members from both the CTRL and STO took a similar westerly track, the STO tracks took a more northwesterly turn overall (as evidenced by the ensemble mean as well) and were slightly slower (Figs. 28a and 29a). Spread increased in the later stages of the forecast. In terms of intensity, there were two forecasts for genesis in the STO forecasts, but there were many fewer vortex-like circulations than in CTRL. On 1 Sep,

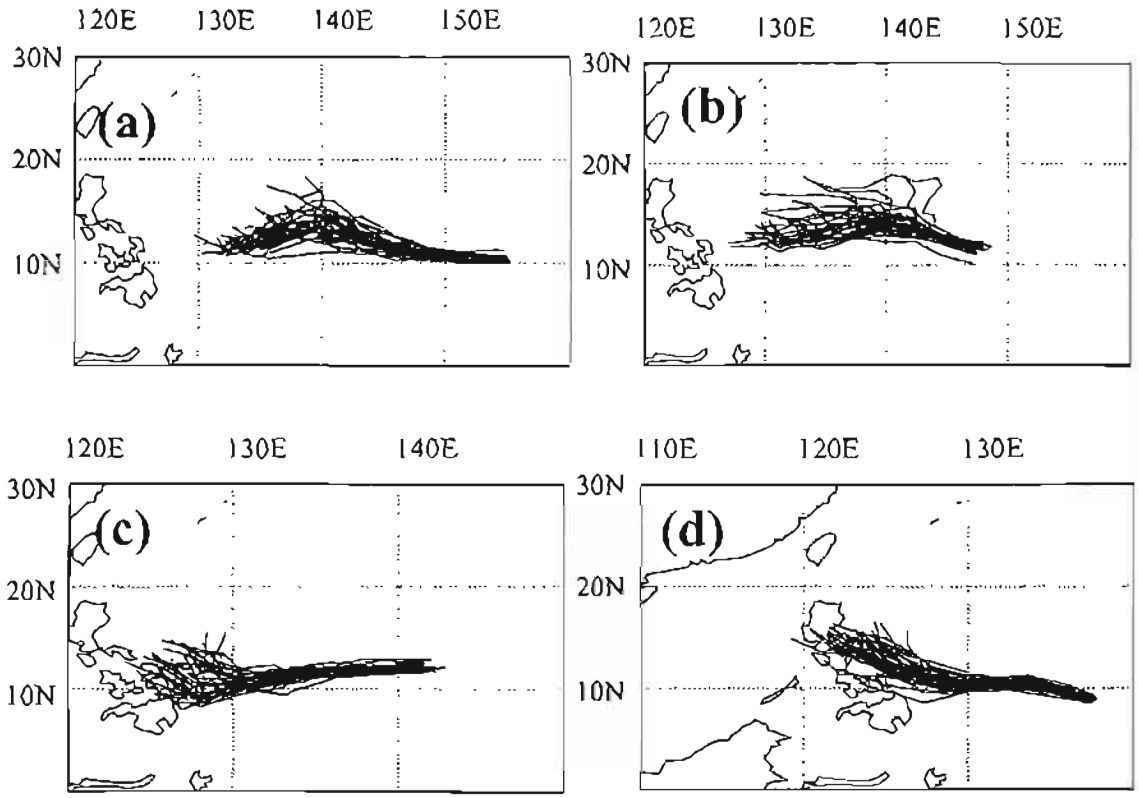


Figure 28: The tracks of 0-120 h CTRL ensemble forecast for TCS030 from 0000 UTC 31 Aug-3 Sep 2008 (corresponding to Fig. 28 a-d, respectively), compared with the JTWC best track.

there were only minor differences in the track forecasts. STO was more divergent toward the end of the forecast period, but spread was not noticeably different from CTRL (Figs. 28b and 29b). There was also a minor shift to the northwest in the STO mean. The major difference occurred in the intensity forecasts, with nine members predicting genesis (compared to zero for CTRL). Even the vortex-like cases tended to show better organization in STO than CTRL. In the forecasts from 2 Sep, the STO tracks were somewhat similar to the CTRL tracks, and did not have a significant increase in spread (Figs. 28c and 29c). The STO mean indicated a turn to the northwest which occurred slightly sooner than in CTRL. Intensity forecasts continued to be stronger in STO, with more cases of both genesis and vortex like development than in CTRL. The primary difference in the track forecasts from 3 Sep was increased spread from STO in the later stages of the forecast (Fig. 29d). Some STO members were faster than any CTRL members. Both ensemble means were nearly identical. While in general most of the STO forecasts predicted a weak circulation, there were two cases of cyclone genesis, which again was more than CTRL.

3.5 Overall evaluation

3.5.1 Predictability: Pregenesis vs. postgenesis

The genesis predictability is assessed here in a similar fashion as it was for the Atlantic cases. Table 7 summarizes these results in terms of the number of members predicting genesis, vortex-like development, dissipation, and nondevelopment for each individual cyclone. In general, the CTRL had low probabilities for genesis (as defined by the criteria in this paper). Averaged for each cyclone, the genesis rate did not exceed 40% for any one individual case. Nuri's intensity forecasts were initially very weak, but in the

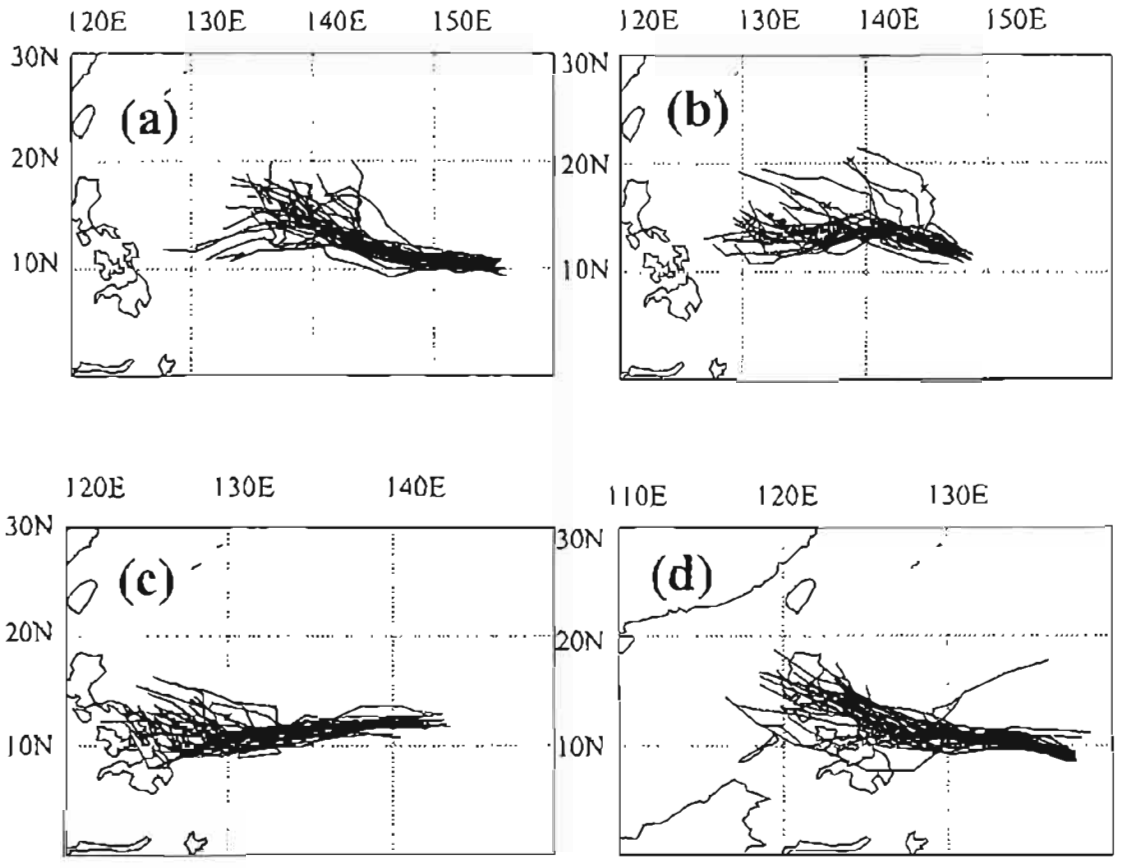


Figure 29: The tracks of 0-120 h STO ensemble forecast for TCS030 from 0000 UTC 31 Aug-3 Sep 2008 (corresponding to Fig. 29 a-d, respectively), compared with the JTWC best track.

Table 7: Predictability of each cyclone for different lead times relative to the system being designated a tropical depression by NHC. Values represent number of members (out of 14) predicting genesis (G), vortex-like development (V), premature dissipation (D), and nondevelopment (N), respectively. Tracking results from CTRL and STO shown separately.

Lead Time	Nuri		Sinlaku		Hagupit		Jangmi		Higos	
(approx. days)	CTRL G/V/D/N	STO G/V/D/N	CTRL G/V/D/N	STO G/V/D/N	CTRL G/V/D/N	STO G/V/D/N	CTRL G/V/D/N	STO G/V/D/N	CTRL G/V/D/N	STO G/V/D/N
-3	0/1/0/32	3/2/0/28	0/9/0/24	1/2/0/30	22/7/2/2	30/2/0/1	0/0/0/33	7/3/0/23	0/0/0/33	2/4/1/26
-2	4/18/0/11	4/3/0/26	1/0/0/32	3/4/0/26	10/7/4/12	22/4/0/7	0/3/0/30	8/2/0/23	8/25/0/0	17/12/0/4
-1	33/0/0/0	25/6/0/2	1/1/0/31	11/4/0/18	4/5/1/23	10/6/0/17	12/9/0/12	18/2/0/13	1/31/0/1	19/10/0/4
0	33/0/0/0	31/0/2/0	33/0/0/0	33/0/0/0	33/0/0/0	28/2/2/1	33/0/0/0	30/3/0/0	23/10/0/0	26/7/0/0
1	33/0/0/0	33/0/0/0	33/0/0/0	33/0/0/0	33/0/0/0	33/0/0/0	33/0/0/0	33/0/0/0	33/0/0/0	33/0/0/0
2	33/0/0/0	33/0/0/0	33/0/0/0	33/0/0/0	33/0/0/0	33/0/0/0	33/0/0/0	33/0/0/0	33/0/0/0	33/0/0/0

final pregenesis forecast, all members predicted development. However, all these forecasts for genesis were toward the end of the forecast period, well west of the actual genesis location (Fig. 16c). Sinlaku had the worst intensity forecasts of any of the cases with an aggregated 2% predictability rate. The forecasts for Hagupit demonstrated an unexpected trend, with each subsequent forecast approaching genesis having lower genesis predictability. With all of the pregenesis forecasts predicting premature recurvature, none of the predicted genesis locations were near the actual cyclone location (Fig. 20a-c). Hagupit also had several cases of premature dissipation, the only case for which this forecast occurred. Jangmi had poor predictability at longer lead times, with zero members predicting genesis at 2- and 3-day lead times and only 12 members with 1-day lead time. Those members were grouped in 3 clusters in terms of genesis location, with a few near the actual genesis location (Fig. 22c). The intensity forecasts for Higos, which only briefly strengthened to tropical storm status, were appropriate with nearly twice the number of members predicting vortex-like criteria than genesis criteria. For all cases and lead times averaged, the genesis rate was only 23%, which is certainly poor for such short lead times.

However, by combining genesis and vortex-like cases as a metric for circulation predictability, the numbers improved somewhat. Nuri, Hagupit, and Higos all surpassed the 50% mark, with Higos surging to near 75%. Sinlaku and Jangmi increased only marginally, and both still failed to reach a 25% predictability rate. By this method, the overall average for all cases doubled, up to a rate of 46%.

The other way to evaluate the genesis predictability is to examine the number of members predicting genesis by lead time. Since the 0000 UTC forecasts had different

relative lead times from tropical depression designation, the model initializations were grouped together as close as possible, with preference toward a particular full day (e.g., 3-day, or 72-h, lead time). The results are summarized in Table 8. As expected, the genesis rate increased as lead time decreased. However, the overall rates were still low, starting with 13% at 3-day lead time and only increased to 31% at 1-day lead time. Combining genesis and vortex-like criteria, the 2-day lead time rate made the most increase, up to 46%, while only the 1-day lead time broke the 50% barrier.

Intensity estimates improved dramatically after actual genesis time. For CTRL forecasts, almost all members met genesis criteria for all postgenesis forecasts. The only exception was Higos, where 10 members met vortex-like criteria only on the day of genesis ("zero lead time"). Technically, though, this forecast occurred 12 (6) h before the system was designated a tropical depression (storm), which would exclude it from other "postgenesis" forecasts. Postgenesis intensity forecasts were reasonably accurate, with fluctuations or weakening only occurring with forecast interactions with land or midlatitude features. However, since there were often erroneous track forecasts that moved the cyclone into these features, the intensity forecasts were correspondingly inaccurate.

For the two nondeveloping systems, the model performed reasonably well in predicting them to remain weak disturbances (Tables 9-10). The forecasts met genesis criteria in only four forecasts of TCS017 and none for TCS030. There were more forecasts for vortex-like development for both systems, but much more so (35 versus 14) for TCS030. This fact is interesting since forecasters on the TCS-08 field experiment had much higher confidence that TCS017 could potentially develop into a cyclone. However,

Table 8: Probability of genesis in ensemble for each lead time (forecast from number of days) in CTRL, combined for all five named cyclones.

Lead Time (days relative to genesis)	Genesis (G)	Vortex-Like (V)	G+V	Nondevelop (N+D)
-3	22/165	17/165	39/165	126/165
-2	23/165	53/165	76/165	89/165
-1	51/165	46/165	97/165	68/165

Table 9: Predictability of TCS017 in 5 day forecasts started from different lead times. Values represent number of members (out of 14) predicting genesis, vortex-like development, premature dissipation, and nondevelopment, respectively. Number of members formatted as: CTRL/STO.

Forecast Date	**19 Aug**	**20 Aug**	**21 Aug**	**22 Aug**
Development	0/1	4/8	0/1	0/0
Vortex-Like	2/4	6/6	4/2	2/4
Dissipation	0/0	0/0	0/0	0/0
Nondevelopment	31/28	23/19	28/30	31/29

Table 10: Predictability of TCS030 in 5-day forecasts started from different lead times. Values represent number of members (out of 14) predicting genesis, vortex-like development, premature dissipation, and nondevelopment, respectively. Number of members formatted as: CTRL/STO.

Forecast Date	31 Aug	1 Sep	2 Sep	3 Sep
Development	0/2	0/9	0/2	0/2
Vortex-Like	18/4	11/11	2/6	4/4
Nondevelopment	0/0	0/0	0/0	0/0
Dissipation	15/27	22/13	31/25	29/27

when distributed over the four forecasts for each case, these false alarm rates were still relatively small. When tallying the members which predicted nondevelopment (there were no cases of premature dissipation), the probability of detection was 86% for TCS017 and 73% for TCS030, for an average of 80%.

3.5.2 Ensembles with stochastic convection vs. without stochastic convection

On average, the STO intensity forecasts were stronger than those of CTRL. This fact was readily apparent in the overall genesis percentage (totaled for all cyclones and all lead times): 39% compared to 23% for CTRL. In addition, it also can be seen in the number of members that predicted cases that met genesis criteria in most cases. However, the opposite was true for Nuri, with only 32% of STO members predicted genesis totaled over all three lead times compared to 37% for CTRL. This reduction was partially due to the fact that only 25 members (compared to all 33 CTRL members) met genesis criteria in the final pregenesis forecast. Predicted genesis locations in STO were much more evenly spread along the projected track than in CTRL (Figs. 16c and 17c). The intensity forecasts for Sinlaku improved slightly in STO, but overall, the genesis percentage was only 15%. Most of the forecasts for genesis were not located near the actual cyclone track (Figs. 19b-c). Similar to CTRL, the number of members predicting the genesis of Hagupit decreased as the actual genesis time approached. However, the total number of members predicting genesis significantly increased over CTRL, with a combined genesis rate of 63% (whereas CTRL had a rate of 36%). The number of STO members predicting the genesis of Jangmi increased over CTRL for all three pregenesis lead times. As shown in Figs. 23a-c, these predicted cyclones were forecast to develop at varying rates, with genesis

locations at various locations along the track for all three lead times. The cyclone-total genesis rate was still comparatively low to the other cases at 33%. An example of the STO strength bias was present in the intensity forecasts for Higos. Whereas the CTRL forecasts had double the number of vortex-like cases as genesis cases, STO had double the number of genesis cases as vortex-like cases. This disparity was represented in the cyclone-total genesis rate, which is 48%. However, these forecasts may be reasonable in predicting genesis, since the small number of cases here limits comparative opportunities. In terms of storm strength, STO had many cases in which members predicted the system to intensify, some comparable to typhoon strength, which was an over-forecast of reality. Most of the forecasts for development were evenly distributed along the track, although they generally occurred after the actual cyclone developed (Figs. 25b-c).

When combining genesis and vortex-like criteria, the Nuri STO intensity forecasts continued to fall short of the number of CTRL members, with only a 43% predictability rate compared to 57% for CTRL. Sinlaku's total predictability rate remained low for STO, at 25%. Combined STO forecasts for Hagupit showed a large increase, up to 75%. Jangmi's total was more moderate, with a 40% predictability rate. The genesis plus vortex-like rate for Higos remained about the same for STO and CTRL, just below 75%. This method (combined circulation predictability for all cyclones and all lead times) produces the highest overall predictability percentage, which was 53%.

In both of the nondeveloping cases, STO proved to have a higher false alarm rate (Tables 9-10). For TCS017, more STO members predicted both genesis and vortex-like criteria than CTRL. On the other hand, TCS030 had more forecasts for genesis but fewer forecasts of vortex-like criteria than CTRL. Averaging the 8% genesis predictability rate

for TCS017 and 11% rate for TCS030 produced an overall STO false alarm rate of 9%, which was higher than the 2% rate for CTRL. The STO probability of detection rates of the nondeveloping systems were similar to CTRL, but were lower for both systems. In turn, the overall STO probability of detection was 75%, which was slightly less than the 80% for CTRL. Statistics from more cases are needed to determine if these false alarm rates are reasonable.

Averaged over each lead time, STO forecasts had higher predictability rates than CTRL for all lead times (Table 11). None of the rates fell below 25% but also did not rise above 50% except for one day lead time. The differences between STO and CTRL were more pronounced for genesis-only than for genesis plus vortex-like criteria. There were also more vortex-like cases as forecast lead time decreased, as evidenced by the differences in the STO predictability rate between genesis-only and genesis-plus-vortex-like criteria. As expected, the highest predictability rate occurred for combined genesis and vortex-like criteria for one day lead time, at 67%.

The STO intensity forecasts were comparatively different from those of CTRL (Table 7). While not overwhelming, the postgenesis intensity initialization varied more for STO. There were more cases of the postgenesis intensity not meeting genesis criteria. For Nuri, there were two cases of premature dissipation from the forecast coinciding with genesis (but 12 h before tropical storm designation). Sinlaku was the only case for which all members met genesis criteria for all the postgenesis forecasts. There were varying results from the Hagupit initialization 6 h after tropical depression formation. In two members, only vortex-like criteria were met, in addition to two member forecasts of premature dissipation, and one member did not predict development at all. There were

Table 11: Probability of genesis in ensemble for each lead time (forecast from number of days) in STO, combined for all five named cyclones.

Lead Time (days relative to genesis)	Genesis (G)	Vortex Like (V)	G+V	Nondevelop (N+D)
-3	43/165	13/165	56/165	109/165
-2	54/165	25/165	79/165	86/165
-1	83/165	28/165	111/165	54/165

three members that only showed vortex-like criteria on the day of Jangmi's genesis (initialized 6 h after tropical depression designation). The earliness of the "genesis day" forecasts for Higos was also evident in STO, with seven members only meeting vortex-like criteria. The STO postgenesis forecast intensities tended to be, in general, stronger than CTRL. While not dramatic in the typhoon cases (since all four became intense typhoons), over-intensification was seen more often in the weaker systems. Comparatively, there were many more forecasts of Higos strengthening dramatically in STO than CTRL. Also, while not "postgenesis," there were a few forecasts for the nondeveloping systems to strengthen significantly that were not seen in CTRL. In addition, like CTRL, STO intensity forecasts often depended on the track forecasts.

3.5.3 Mean, spread, and accuracy of track forecasts

Table 12 summarizes the values for the ensemble mean error (compared to the best track data) and the standard deviation of the ensemble track, averaged for all forecast times (0-120 h). By far, Hagupit had the largest track errors in the pregenesis forecasts. In the cases of Nuri, Sinlaku, and Higos, the mean errors increased close to genesis time and then decreased. These were the cases with the forecasts for early recurvature, which became less erroneous as the system developed. The ensemble mean errors followed a similar pattern in both CTRL and STO (and was also seen in the track plots), with neither seemingly having an overwhelming advantage. There were no noticeable trends in the error standard deviation values. In general, Hagupit and Higos had the highest standard deviation values in both CTRL and STO. For the numerical track standard deviations, there did not appear to be a remarkable difference between CTRL and STO. The CTRL was higher in some cases, while STO was higher in others.

Table 12: Ensemble mean error and spread (km) for all cyclones. Lead time denotes the forecast from the number of days before the system was designated a tropical depression. CTRL and STO shown separately.

Lead Time	Nuri		Sinlaku		Hagupit		Jangmi		Higos	
(days)	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
-3	181.2	40.55	155.9	123.99	1173.5	212.03	605.6	134.07	421.1	337.01
-2	216.7	78.31	338.8	73.51	1411.6	482.76	490.4	164.57	566.6	236.35
-1	560.3	391.73	306.6	169.76	1038.6	509.36	422.2	253.41	673.0	532.34
0	393.3	217.97	414.6	507.3	498.3	344.1	262.8	260.08	577.7	401.62
1	314.8	309.24	273.0	313.66	458.5	383.55	317.8	302.13	350.6	134.25
2	340.0	370.74	138.6	104.02	197.1	171.53	364.0	294.41	109.2	81.77
	STO Mean	STO SD	STO Mean	STO SD	STO Mean	STO SD	STO Mean	STO SD	STO Mean	STO SD
-3	141.7	47.39	147.8	121.78	1118.4	256.07	543.4	42.2	443.0	289.85
-2	309.8	229.73	285.2	120.3	1481.5	617.32	484.9	138.45	551.8	398.3
-1	562.8	358.39	279.6	267.01	1031.2	586.75	340.4	187.9	588.5	447.42
0	330.0	241.69	323.5	379.85	526.7	529.99	211.1	199.5	546.1	315.03
1	259.9	251.49	175.3	189.34	459.8	389.07	280.7	293.8	311.8	135.77
2	300.5	286.24	130.6	88.19	195.1	184.73	336.2	283.59	100.2	80.94

While the averaged error standard deviations discussed above change little between CTRL and STO, STO track forecast spread increased over CTRL, as seen in the track forecasts. Even though the tracks of "member 0" were sometimes different between CTRL and STO, the ensemble means were fairly similar. Thus, these cases have shown that the stochastic convection scheme served to enhance the amount of spread in the ensemble forecasts without dramatically changing the nature of the forecasts themselves. However, despite the increase in spread, STO forecasts added limited value when overall model skill was poor. For example, there were many cases of recurvature that did not verify, and overall, the STO forecasts were similar to the CTRL forecasts.

Similar to the Atlantic cases, the ensemble mean and standard deviation of the track have been compared in a scatter plot (Fig. 30). A linear fit has been applied to depict the correlation between these two variables. In addition, the cases have been divided into pregenesis and postgenesis forecasts. The coefficient of determination (square of the correlation coefficient) is used as a metric for correlation again. Like the Atlantic cases, the correlation in the pregenesis phase was low, with CTRL and STO on either side of a 0.50 R^2 value. Fig. 12 also shows that both CTRL and STO were under-dispersive in the pregenesis forecasts (the spread was small than the mean error). Also in a similar fashion to the Atlantic cases, the coefficient of determination improved for the postgenesis forecasts, where CTRL increased marginally to 52%. STO showed dramatic improved, with almost 80% of the variance explained. The postgenesis STO forecasts made the best case to justify that the ensemble spread represents the uncertainty in the ensemble mean. While STO was promising in this aspect, R^2 values were not significant for all of the pregenesis forecasts and postgenesis for CTRL. It should be noted that the

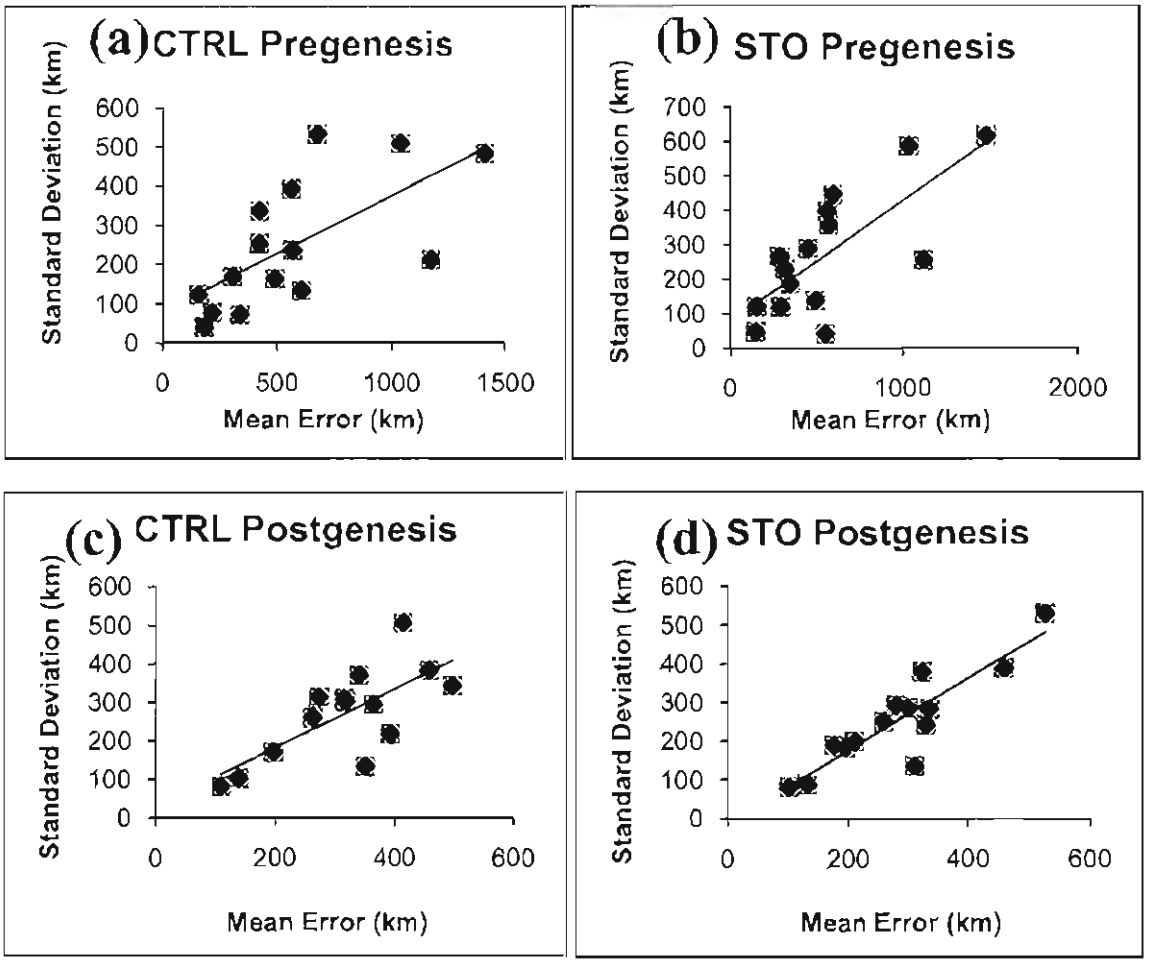


Figure 30: Scatter plots of ensemble mean error (km) vs. ensemble spread (represented by standard deviations of the ensemble members relative to the ensemble mean in km) for each forecast period (1 to 5 days) for all cyclones. Forecasts are divided into two groups by the forecast lead time in (a and c) pregenesis, and (b and d) postgenesis. CTRL is shown in panels a and b, while STO is shown in panels c and d.

correlation between ensemble standard deviation and track errors was higher in STO than in CTRL in both pregenesis and postgenesis forecasts. These increased values show that STO is more valuable than CTRL in representing ensemble uncertainties, as well.

3.6 Summary

In this chapter, seven cases in the Western Pacific Basin during Aug and Sep 2008 were evaluated. There were four typhoons, one tropical storm, and two nondeveloping systems. The NOGAPS ensemble system was evaluated in its ability to predict the formation and development of these tropical systems. In addition, the control ensemble system was compared in its base form against a newer version that included stochastic convection.

The ensemble track forecasts had various results. Three of the cases had consistent forecasts for premature recurvature, with few (if any) members suggesting courses that were close to the actual track. Stochastic convection had little effect on these erroneous tracks, suggesting the issue may have been with the base NOGAPS model. There was not a noticeable change in the accuracy between pregenesis and postgenesis tracks. In fact, some of the erroneous forecasts for recurvature occurred after the cyclone had developed. However, there were no changes to the model after genesis occurred (such as the vortex relocation in the NCEP ensemble). Spread in the track forecasts was generally small. Therefore, in terms of the track forecasts, the ensemble offered little additional value over a single deterministic forecast.

Pregenesis intensity forecasts were also poor. Except for Hagupit, all of the named cyclones had genesis predictability less than 50% at 3-day lead time. The genesis predictability improved with decreasing lead time (again, except for Hagupit), but

Sinlaku was the only case where a high percentage of members forecasted genesis criteria one day in advance. Most notably, however, the intensity forecasts (or at the minimum, the initialization) improved dramatically after genesis. Except in a few cases, nearly all of the ensemble members initialized the cyclones meeting genesis criteria, and maintained their intensity until the forecasts called for extratropical transition or landfall. In general, the false alarm rates were small for the nondeveloping systems.

The primary motive behind including stochastic convection in the ensemble was to increase the ensemble spread in the tropics. It definitely succeeded in this respect. In nearly all of the cases, the stochastic ensemble had increased track spread over the control ensemble. The ensemble means were similar for both the stochastic and control ensembles, suggesting this increase in spread was achieved without sacrificing other components of the forecast. However, this asset also served to limit the forecasts when the base model was incorrect (e.g., for cases with premature recurvature), and possibly prevented more members from encompassing the observed cyclone track. In addition to the increased spread in tracks, the stochastic ensemble also had improved pregenesis intensity forecasts. Overall, genesis predictability was higher in the stochastic ensemble than the control ensemble. As a negative side effect, though, this tendency for an increase in intensity also was evident in weaker systems. For Higos (only a tropical storm) and both nondeveloping systems, there were more stochastic members predicting genesis (and in some cases, significant intensification) than control members. Whether these forecasts are physically reasonable, or if they would be consistent over many situations, could not be determined with the limited number of cases here.

When determining if the ensemble spread is a good indicator of the uncertainty in the forecast error, the postgenesis forecasts showed a better relationship than the pregenesis forecasts. In addition, the stochastic ensemble showed a higher correlation between the size of ensemble spread and track errors than the control ensemble. The stochastic forecasts in the postgenesis phase demonstrated the most convincing correlation between the ensemble mean error and spread, with the standard deviation explaining 80% of the variance of the average error.

CHAPTER 4

CONCLUDING REMARKS

4.1 Summary

The primary objective of this thesis was the evaluation of the performance of global ensemble prediction systems in the tropics. Specifically, interest was focused on the formation, evolution, intensity, and track forecasts of tropical cyclones. Ensemble skill in forecasting nondeveloping tropical systems was also evaluated. The motivation behind this work involved a lack of previous studies of single model ensembles. Since tropical forecasters generally rely on multimodel ensembles (for various reasons), little attention has been given to the performance of single model ensembles in the tropics. To accomplish this goal, two sets of cases in two oceanic basins were studied independently using two global operational ensembles. Each case was tracked in the 5-day forecasts, including ensemble initializations from several days before to several days after genesis (if genesis occurred).

Firstly, the NCEP global ensemble was evaluated in the Atlantic Basin. Since a standard tracking method has not previously been defined, the development of an effective tracking method was necessary. Several atmospheric variables and pressure levels were evaluated in their effectiveness and accuracy of tracking tropical cyclones in both their pre and postgenesis phases. A combination of vorticity, wind vectors, and

geopotential height at 850 hPa was adopted to subjectively track the systems. This methodology proved to have a high level of consistency and minimized errors in most cases.

In the Atlantic basin, seven cases were evaluated: four hurricanes, one tropical storm, and two nondeveloping waves. For the intensity forecasts, the systems with better organization were predicted more accurately. Debby (in later stages of development), Florence, and Helene all had reasonable genesis predictability in the pregenesis forecasts. Meanwhile, the ensemble did not predict the formation of Ernesto and Gordon. Both of these cyclones were initially weaker and smaller in scale. They developed much farther west in the Atlantic as well. Combining all cyclones, development was predicted about 70% of the time in pregenesis forecasts. When separated by lead time, shorter lead times were found to have higher genesis rates. Postgenesis intensity forecasts were somewhat improved, although there were many cases where the initializations and subsequent forecasts were too weak. For the two nondeveloping systems, the ensemble performed well by not predicting false alarms for genesis. Track forecasts were similar to intensity in that they improved with decreased lead time, and were also more accurate for the stronger cyclones. However, in all cases, the tracks dramatically improved after the systems became tropical depressions and the vortex relocation scheme was implemented. The operational GFS with a higher resolution proved to be more likely to predict cyclone genesis than the ensemble members. However, the probabilistic component of the ensemble proved to be equally skillful, with a majority of the members usually falling into the same intensity category as the GFS. The spread-skill relationship proved to be

strong for these cases. In postgenesis forecasts, 83% of the variance in the ensemble mean error is explained by the ensemble spread.

Seven cases were also evaluated in the West Pacific Basin. Similarly, there were four typhoons, one tropical storm, and two nondeveloping systems. These cases were tracked with the NOGAPS ensemble system. Overall, track forecasts lacked accuracy, with three cases demonstrating several forecasts of premature recurvature. There were limited relationships between track accuracy and system strength or lead time. Although in general, the tracks became more accurate with shorter lead times, there were no distinguishing factors between pre and postgenesis forecasts. Pregenesis intensity forecasts were also inaccurate, with most of the cyclone cases having well below 50% genesis predictability. Postgenesis forecasts were much improved, with nearly all members initializing the cyclones with the correct intensity and maintaining the intensity through an appropriate forecast duration. False alarm rates were small for the nondeveloping systems.

The NOGAPS ensemble was also run with a stochastic convection scheme for the seven West Pacific cases. This scheme is supposed to increase the ensemble spread to a more appropriate level in the tropics. While track spread increased in nearly all the forecasts, the tracks were still constrained by similarly erroneous ensemble means and control members. Pregenesis intensity forecasts were also improved, with the stochastic ensemble having much higher genesis predictability. However, the stochastic ensemble also tended to over-intensify weaker waves. In terms of the spread-skill relationship, the postgenesis stochastic forecasts showed the highest correlation, with 80% of the ensemble mean error variance explained by the track standard deviation. The other

configurations (pregenesis stochastic; pre- and postgenesis control) of the ensemble showed much weaker correlations.

4.2 Discussion

Several aspects of the ensemble prediction systems were investigated in this thesis. In the majority of the cases, the ensemble mean track was somewhat similar to the control track. However, there were many more examples of deviation between those two tracks in the NCEP cases than the NOGAPS cases. Although there did not appear to be any relationship between the ensemble spread and forecast time relative to genesis, there was often more track spread for weaker systems due to erratic tracks. The standard deviation in the track forecasts was a more significant indicator of the forecast uncertainty (error) in postgenesis forecasts only. In addition, this metric was stronger in the pregenesis forecasts for the NCEP ensemble than the NOGAPS ensemble. The NCEP ensemble included a vortex relocation scheme. Although there were only five developed cases, the vortex relocation appeared to have a positive effect on the track forecasts. With the improvements seen in this study, a vortex relocation scheme would be a worthwhile consideration for other models. The NOGAPS ensemble was compared in its base state and when stochastic convection was added. Stochastic convection succeeded at improving the ensemble spread in the tropics, thus making it a successful addition to the ensemble. However, track spread was still small in some cases, and the ensemble skill overall was limited by erroneous model forecasts. Stochastic convection may also have a negative impact on intensity forecasts, with several cases of weak systems being forecasted to strengthen dramatically.

Even though two different models were studied here, it is difficult to directly compare them since the cases are completely independent. Trends seen in the models are also challenging to assess because the cases occurred in two different basins, where issues such as weather patterns and data availability may cause the models to perform differently. While the NCEP ensemble was not overly impressive (especially for Ernesto and Gordon), it fared much better in track and intensity forecasts than the NOGAPS. The NOGAPS ensemble had consistent issues with premature recurvature. Pregenesis intensity forecasts tended to be too weak in the NOGAPS. On the other hand, the NCEP ensemble did not appear to have systematic biases based on the cases studied here. The NCEP ensemble system has been in use longer than the NOGAPS ensemble, meaning further development has been possible to prevent similar issues. Goerss and Reynolds (2008) found that the NOGAPS ensemble was more skillful than the NCEP ensemble for some Atlantic cases, so the evidence presented here is case-dependent only.

The limited number of case studies in this thesis presented mixed results regarding the usefulness of a single model ensemble in tropical cyclone prediction. Some cases, such as Helene, showed that the ensemble is a worthwhile tool in predicting both track and intensity. In other cases, such as the forecast initialized 42 h before the genesis of Ernesto, the ensemble mean can be much more accurate when the deterministic GFS forecast happened to be an outlier. However, the ensemble performed poorly in a number of cases. Both track and intensity forecasts were problematic in the pregenesis stages of Ernesto and Gordon. Nuri, Sinlaku, Hagupit, and Higos all had forecasts of erroneous track curvature, occurring both before and after genesis. However, this trend may be an overall indictment of bias in the NOGAPS model rather than individual issues with the

ensemble or particular cases. Overall, though, nothing presented in this thesis would rule out the proposition of including a single model ensemble as a tool in the forecaster's "toolbox." Most cases showed that the ensemble means (both in track and intensity) were at least equally adequate to the deterministic forecasts of the respective models studied. Therefore, one of the major conclusions of this research is that a single model operational ensemble can be a valuable prediction tool in the tropics.

While the ensemble system may be a useful tool, there are some limiting factors to consider. As seen in several cases here, particularly the NOGAPS, the ensemble may not account for an accurate error range. However, the ensemble envelope, regardless of its size, still represents a suite of possibilities to the forecaster. The amount of spread may be considered to assess the confidence of the forecast, but forecasters would be wise to evaluate other independent model (and ensemble) forecasts before making conclusions about the tropical cyclone's expected track and intensity. Another factor to keep in mind is that the ensemble forecasts are only as good as the base model's performance. Improvements to the base model (NOGAPS and GFS in this case) are needed to improve the forecasts of the ensemble. It is also possible more sophisticated perturbation methods are needed, particularly some designed to accurately represent uncertainty in the tropics. Finally, at least in these cases, there was no clear advantage of having a 32-member ensemble over a 14-member ensemble. In addition, the higher resolution deterministic GFS proved to be more accurate for intensity forecasts. Therefore, it may be more productive to focus ensemble improvements on increasing resolution instead of increasing the number of members.

4.3 Suggestions for future work

This thesis suggested single model ensembles can be a worthy tool to be utilized in tropical prediction. However, with only 10 developed cyclones and four nondeveloped systems, it must be noted that the sample size of the evaluation is small. More case studies need to be done in the future to fully evaluate the skill of ensembles in predicting tropical cyclone development. In order to enhance the skill of ensemble forecasting, it is necessary to evaluate ensemble forecasting itself, particularly the impact of the size and distribution of the initial perturbations on the ensemble skill of forecasting tropical cyclone development. Also, by assessing the spread of the initial perturbations, conclusions could be drawn on how they affect the spread and skill of the subsequent forecasts. These investigations could be particularly useful when determining the differences between the control and stochastic NOGAPS ensembles.

One key aspect that was not investigated here is the role of atmospheric variables in the forecast and how the ensemble initialized them. In the local atmosphere, factors such as ambient vorticity, temperature, and wind shear can be crucial to tropical cyclone formation. On the large scale, strength and placement of troughs and ridges are critical to tropical cyclone movement and occasionally strengthening (e.g., favorable upper level divergence). To expand this research, it would be interesting to determine the mechanisms for the poor forecasts by conducting sensitivity tests. Understanding these factors and their uncertainty could help improve both the ensemble system as well as the tropical cyclone forecasts.

Finally, it may be useful to follow the spirit of the Goerss and Reynolds (2008) research. To truly evaluate the effectiveness of model forecasts in a relative sense, it is

important to conduct inter-comparisons between models. In addition to the NCEP GFS ensemble, the ECMWF ensemble is widely respected amongst forecasters and thus would make an intriguing candidate for study in both the Atlantic and West Pacific cases. The key issue to tackle is how these single model ensembles compare directly to the multimodel (consensus) forecast tools in the tropics. This type of proof would be necessary to convince forecasters that single model ensembles can be as skillful as consensus forecast tools in the tropics. In addition, it would be useful to investigate how single model and multimodel ensembles may be used together to produce a superior forecasting tool. Comparing ensemble prediction systems in other ocean basins could also help complete a full assessment of ensemble prediction in the tropics.

REFERENCES

Aberson, S.D., 1999: Ensemble-based Products to Improve Tropical Cyclone Forecasting. Preprints, *23rd Conf. on Hurricanes and Tropical Meteorology*, Dallas, TX, Amer. Meteor. Soc., 843-844.

Aberson, S.D., 2001: The Ensemble of Tropical Cyclone Track Forecasting Models in the North Atlantic Basin (1976–2000). *Bull. Amer. Meteor. Soc.*, **82**, 1895–1904.

Aberson, S.D., M.A. Bender, and R.E. Tuleya, 1998: Ensemble forecasting of tropical cyclone tracks. Preprints, *12th Conf. on Numerical Weather Prediction*, Phoenix, AZ, Amer. Meteor. Soc., 290-292.

Albignat, J.P., and R.J. Reed, 1980: The Origin of African Wave Disturbances during Phase III of GATE. *Mon. Wea. Rev.*, **108**, 1827–1839.

Beven, J. L., 1999: The boguscane—A serious problem with the NCEP medium range forecast model in the Tropics. Preprints, *23d Conf. on Hurricanes and Tropical Meteorology,* Dallas, TX, Amer. Meteor. Soc., 845–848.

Beven, J.L., 2006: Tropical Cyclone Report: Hurricane Florence. National Hurricane Center. Available http://www.nhc.noaa.gov/2006atlan.shtml.

Bishop, C.H., and Z. Toth, 1999: Ensemble Transformation and Adaptive Observations. *J. Atmos. Sci.*, **56**, 1748–1765.

Blake, E.S., 2006: Tropical Cyclone Report: Hurricane Gordon. National Hurricane Center. Available http://www.nhc.noaa.gov/2006atlan.shtml.

Brown, D.P., 2006: Tropical Cyclone Report: Hurricane Helene. National Hurricane Center. Available http://www.nhc.noaa.gov/2006atlan.shtml.

Buizza, R., P.L. Houtekamer, Z. Toth, G. Pellerin, M. Wei, and Y. Zhu, 2005: A Comparison of the ECMWF, MSC, and NCEP Global Ensemble Prediction Systems. *Mon. Wea. Rev.*, **133**, 1076–1097.

Burpee, R., 1972: The Origin and Structure of Easterly Waves in the Lower Troposphere of North Africa. *J. Atmos. Sci.*, **29**, 77–90.

Chen, T.C., 2006: Characteristics of African Easterly Waves Depicted by ECMWF Reanalyses for 1991–2000. *Mon. Wea. Rev.*, **134**, 3539–3566.

Cheung, K.K.W., and J.C.L. Chan, 1999: Ensemble Forecasting of Tropical Cyclone Motion Using a Barotropic Model. Part II: Perturbations of the Vortex. *Mon. Wea. Rev.*, **127**, 2617–2640.

Cheung, K.K.W., and R.L. Elsberry, 2002: Tropical Cyclone Formations over the Western North Pacific in the Navy Operational Global Atmospheric Prediction System Forecasts. *Wea. Forecasting*, **17**, 800–820.

Fink, A.H., D.G. Vincent, P.M. Reiner, and P. Speth, 2004: Mean State and Wave Disturbances during Phases I, II, and III of GATE Based on ERA-40. *Mon. Wea. Rev.*, **132**, 1661–1683.

Franklin, J.L., 2006: Tropical Cyclone Report: Tropical Storm Debby. National Hurricane Center. Available http://www.nhc.noaa.gov/2006atlan.shtml.

Fujiwhara, S., 1923: On the Growth and Decay of Vortical Systems. *Quart. J. Roy. Meteor. Soc.*, **49**, 75-104.

Goerss, J.S., 2000: Tropical Cyclone Track Forecasts Using an Ensemble of Dynamical Models. *Mon. Wea. Rev.*, **128**, 1187–1193.

Goerss, J.S., and C.A. Reynolds, 2008: Impact of Stochastic Cumulus on the NOGAPS ET Ensemble Forecasting System. Part II: Tropical Cyclone Track Forecast Performance. Preprints, *28th Conf. on Hurricanes and Tropical Meteorology,* Orlando, FL, Amer. Meteor. Soc.

Hennon, C.C., and J.S. Hobgood, 2003: Forecasting Tropical Cyclogenesis over the Atlantic Basin Using Large-Scale Data. *Mon. Wea. Rev.*, **131**, 2927–2940.

Hogan, T.F., and T.E. Rosmond, 1991: The Description of the Navy Operational Global Atmospheric Prediction System's Spectral Forecast Model. *Mon. Wea. Rev.*, **119**, 1786–1815.

Kerns, B., K. Greene, and E. Zipser, 2008: Four Years of Tropical ERA-40 Vorticity Maxima Tracks. Part I: Climatology and Vertical Vorticity Structure. *Mon. Wea. Rev.*, **136**, 4301–4319.

Knabb, R.D., and M. Mainelli, 2006: Tropical Cyclone Report: Hurricane Ernesto. National Hurricane Center. Available http://www.nhc.noaa.gov/2006atlan.shtml.

Krishnamurti, T.N., C.M. Kishtawal, Z. Zhang, T. LaRow, D. Bachiochi, E. Williford, S. Gadgil, and S. Surendran, 2000: Multimodel Ensemble Forecasts for Weather and Seasonal Climate. *J. Climate*, **13**, 4196–4216.

Liu, Q., S. Lord, N. Surgi, Y. Zhu, R. Wobus, Z. Toth, and T. Marchok, 2006: Hurricane Relocation in Global Ensemble Forecast System. Preprints, *27th Conf. on Hurricanes and Tropical Meteorology,* Monterey, CA, Amer. Meteor. Soc.. April 24-28, 2006, Monterey, CA.

Mackey, B.P., and T.N. Krishnamurti, 2001: Ensemble Forecast of a Typhoon Flood Event. *Wea. Forecasting*, **16**, 399–415.

Marchok, T.P., 2002: How the NCEP Tropical Cyclone Tracker Works. Preprints, *25th Conf. on Hurricanes and Tropical Meteorology*, San Diego, CA, Amer. Meteor. Soc., P1.13.

McBride, J.L., and R. Zehr, 1981: Observational Analysis of Tropical Cyclone Formation. Part II: Comparison of Non-Developing versus Developing Systems. *J. Atmos. Sci.*, **38**, 1132–1151.

McLay, J.G., C.H. Bishop, and C.A. Reynolds, 2008: Evaluation of the Ensemble Transform Analysis Perturbation Scheme at NRL. *Mon. Wea. Rev.*, **136**, 1093–1108.

Pasch, R.J., J.-G. Jiing, F. M. Horsfall, H.-L. Pan, and N. Surgi, 2002: Forecasting tropical cyclogenesis in the NCEP global model. Preprints, *25th Conf. on Hurricanes and Tropical Meteorology,* San Diego, CA, Amer. Meteor. Soc., 178–179.

Perrone, T.J., and P.R. Lowe, 1986: A Statistically Derived Prediction Procedure for Tropical Storm Formation. *Mon. Wea. Rev.*, **114**, 165–177.

Pytharoulis, I., and C. Thorncroft, 1999: The Low-Level Structure of African Easterly Waves in 1995. *Mon. Wea. Rev.*, **127**, 2266–2280.

Reed, R., A. Hollingsworth, W. Heckley, and F. Delsol, 1988: An Evaluation of the Performance of the ECMWF Operational System in Analyzing and Forecasting Easterly Wave Disturbances over Africa and the Tropical Atlantic. *Mon. Wea. Rev.*, **116**, 824–865.

Rennick, M.A., 1999: Performance of the Navy's Tropical Cyclone Prediction Model in the Western North Pacific Basin during 1996. *Wea. Forecasting*, **14**, 297–305.

Reynolds, C.A., J. Teixeira, and J.G. McLay, 2008: Impact of Stochastic Convection on the Ensemble Transform. *Mon. Wea. Rev.*, **136**, 4517–4526.

Rogers, R., S. Aberson, M. Black, P. Black, J. Cione, P. Dodge, J. Dunion, J. Gamache, J. Kaplan, M. Powell, N. Shay, N. Surgi, and E. Uhlhorn, 2006: The Intensity Forecasting Experiment: A NOAA Multiyear Field Program for Improving Tropical Cyclone Intensity Forecasts. *Bull. Amer. Meteor. Soc.*, **87**, 1523–1537.

Teixeira, J., and C.A. Reynolds, 2008: Stochastic Nature of Physical Parameterizations in Ensemble Prediction: A Stochastic Convection Approach. *Mon. Wea. Rev.*, **136**, 483–496.

Thorncroft, C., and K. Hodges, 2001: African Easterly Wave Variability and Its Relationship to Atlantic Tropical Cyclone Activity. *J. Climate*, **14**, 1166–1179.

Toth, Z., and E. Kalnay, 1997: Ensemble Forecasting at NCEP and the Breeding Method. *Mon. Wea. Rev.*, **125**, 3297–3319.

Weber, H.C., 2003: Hurricane Track Prediction Using a Statistical Ensemble of Numerical Models. *Mon. Wea. Rev.*, **131**, 749–770.

Wei, M., Z. Toth, R. Wobus, and Y. Zhu, 2008: Initial Perturbations Based on the Ensemble Transform (ET) Technique in the NCEP Global Operational Forecast System. *Tellus*, **60**, 62-79.

Zawislak, J.A., and E.J. Zipser, 2009: Observations of Seven African Easterly Waves in the East Atlantic During 2006. *J. Atmos. Sci.*: In Press.

Zhang, Z., and T.N. Krishnamurti, 1997: Ensemble Forecasting of Hurricane Tracks. *Bull. Amer. Meteor. Soc.*, **78**, 2785–2795.

Zipser, E. J., Twohy, C. H., Tsay, S. C., Thornhill, K. L., Tanelli, S., et al., 2009: The Saharan Air Layer and the Fate of African Easterly Waves -- NASA's AMMA 2006 Field Program to Study Tropical Cyclogenesis. *Bull. Amer. Meteor. Soc.*, **90**, 1137-1156.